Convention Record of the I·R·E

1954 NATIONAL CONVENTION

Part 5—
Aeronautical Electronics and Telemetry

SESSIONS ON . . .

Aeronautical and Navigational Electronics I

Radio Telemetry and Remote Control I—Systems and Elements

Aeronautical and Navigational Electronics II

Radio Telemetry and Remote Control II—Telemetry

Aeronautical and Navigational Electronics III

Radio Telemetry and Remote Control III—Remote Control

SPONSORED BY

IRE PROFESSIONAL GROUPS ON . . .

Aeronautical and Navigational Electronics

Radio Telemetry and Remote Control

Presented at the IRE National Convention, New York, N.Y., March 22-25, 1954

The Institute of Radio Engineers

Additional Copies

Additional copies of 1954 Convention Record Parts may be purchased from The Institute of Radio Engineers, 1 East 79 Street, New York 21, N. Y., at the prices listed below.

Part	Title	Sponsoring Groups	Prices for Members (M), Libraries (L), & Nonmembers (NM)		
			M	L	NM
1	Antennas & Propagation	Antennas & Propagation	1.25	3.00	3.75
2	Circuit Theory	Circuit Theory	1.25	3.00	3.75
3	Electron Devices & Component Parts	Electron Devices Component Parts	1.50	3.60	4.50
4	Electronic Computers & Information Theory	Electronic Computers Information Theory	1.50	3.60	4.50
5	Aeronautical Electronics & Telemetry	Aeronautical & Navigational Electronics Radio Telemetry & Remote Control	1.50	3.60	4.50
6	Audio & Ultrasonics	Audio Ultrasonics Engineering	1.50	3.60	4.50
7	Broadcasting & Television	Broadcast Transmission Systems Broadcast & Television Receivers	1.50	3.60	4.50
8	Communications & Microwave	Communication Systems Microwave Theory & Techniques Vehicular Communications	1.50	3.60	4.50
9	Medical & Nuclear Electronics	Medical Electronics Nuclear Science	1.50	3.60	4.50
10	Instrumentation & Industrial Electronics	Instrumentation Industrial Electronics	1.25	3.00	3.75
11	Engineering Management & Quality Control	Engineering Management Quality Control	1.00	2.40	3.00
			15.25	36.60	45.75

CONVENTION RECORD OF THE I.R.E.

1954 NATIONAL CONVENTION

PART 5 - AERONAUTICAL ELECTRONICS AND TELEMETRY

TABLE OF CONTENTS

Session 45: Radio Telemetry and Remote Control III - Remote Control
(Sponsored by the Professional Group on Radio Telemetry and Remote Control.)

AN IMPULSE GENERATOR FOR RECEIVER PERFORMANCE MEASUREMENT

By
Joseph H. Vogelman
Chief, Electronic Warfare Laboratory
Rome Air Development Center
Rome, New York

SUMMARY

An impulse generator has been developed consisting of a uniform transmission line which is periodically charging and discharging into a specified output impedance. The ratio of the transmission line impedance and the load impedance being selected so as to produce the flattest possible frequency spectrum over a frequency band from the very low frequency to in excess of 15,000 mc/s. The spectral analysis of the current in the output wave form of the idealized impulse generator shows a flat response to about 85% of resonant frequency over the transmission line. The general design will be considered with particular emphasis on sources of instability and calibration error. The effectiveness of this type of signal generator for measuring the performance of receiving circuits with a pseudo rectangular bandpass with arbitrary phasing will be discussed. It will be shown that the resultant pulse amplitude passing through the receiver will be modified by the bandwidth characteristics of the receiver itself. The case of linear receivers and receivers with phase discontinuities will be considered. The results obtained by impulse measurement of receiver performance will be compared to other means now being utilized.

INTRODUCTION

The problem of measuring radar receiver performance without removing the receiver from operating status has concerned the Armed Services for many years. To do this effectively in field service it was essential that the measuring equipment be as simple as possible and require only a very limited amount of technical skill on the part of the operator. The use of signal generators of the usual pulse modulated rf types proved impractical for several reasons: The need for accurate frequency adjustment, pulse width standardization, and precision attenuators made the equipment technically difficult for the field personnel to operate. In addition, the variability in reading minimum discernible signal or equivalent led to a great dispersion in the readings obtained both with the same operator and between operators. Investigations of the magnitude of the standard deviation for various types of minimum discernible signal resulted in the conclusion that the expected range of measurement dispersion was as large as the allowable degradation in the receiver performance.

To overcome the deficiency of signal generator technique, various types of noise sources have been developed which reduce the complexity of operation to the measurement of a simple dc parameter (or temperature). The noise source of the temperature limited diode or fluorescent type do not have sufficient output to allow them to be fed through a directional coupler into an operating receiver to give an indication at the output of the receiver without elaborate detecting means. The Rome Air Development Center, in cooperation with Professor R.H. George of Purdue Research Foundation, developed a microwave impulse generator designed to overcome the output power deficiencies of the random noise sources.

BASIC CONSIDERATIONS

The intention of the investigation was to produce a device which would generate periodically a large amount of energy in such a manner that its spectral distribution would cover a range of frequencies several magnitudes larger than the receiver bandwidth, and which was essentially uniform over this range. In addition, the magnitude of output of the device had to be directly related to a dc component which could be readily measured. This device would then have the following characteristics: a. It must require no tuning. b. Its output must be readily controlled and readily monitored with a simple dc meter. c. It must give an indication on the output of a radar receiver of sufficient magnitude that it would be readily discernible far above the noise level of the receiver. A series of devices known as microwave impulse generators of various types resulted and it is proposed to consider only the general operating characteristics rather than the details of these equipments.

THEORY OF OPERATION

When an open ended ideal transmission line of characteristic impedance Z_0 is charged to a potential voltage V and short-circuited through a resistance R, current and voltage waves will travel along the transmission line being partially reflected at the resistance and totally

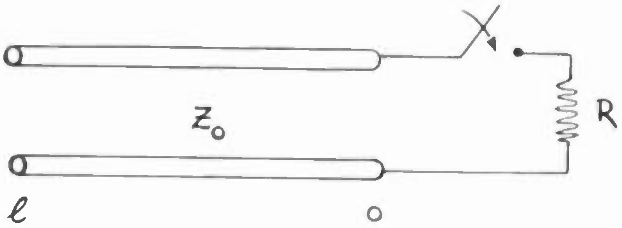


Fig. 1 - Idealized impulse generator.

reflected at the open end. (Figure 1) By investigating the current and voltage relationships for such an ideal transmission line (R= G= 0), the requirements for an idealized impulse noise generator can be readily obtained. We can write the differential equations relating the currents, voltages and characteristics of the transmission line and load resistance and the boundary conditions for the problem as follows:

$$\frac{\partial^2 e}{\partial x^2} - LC\frac{\partial e^2}{\partial t^2} = 0$$

$$\frac{\partial^2 i}{\partial x^2} - LC\frac{\partial^2 i}{\partial t^2} = 0$$

BOUNDARY CONDITIONS:

$e(x,0) = V$ $\quad e'(x,0) = 0$ $\quad i(l,t) = 0$

$i(x,0) = 0$ $\quad i'(x,0) = 0$ $\quad e(0,t) = -Ri(0,t)$

L = LOOP INDUCTANCE IN HENRIES/METER
C = LOOP CAPACITANCE IN FARADS/METER
$Z_o = \sqrt{L/C}$

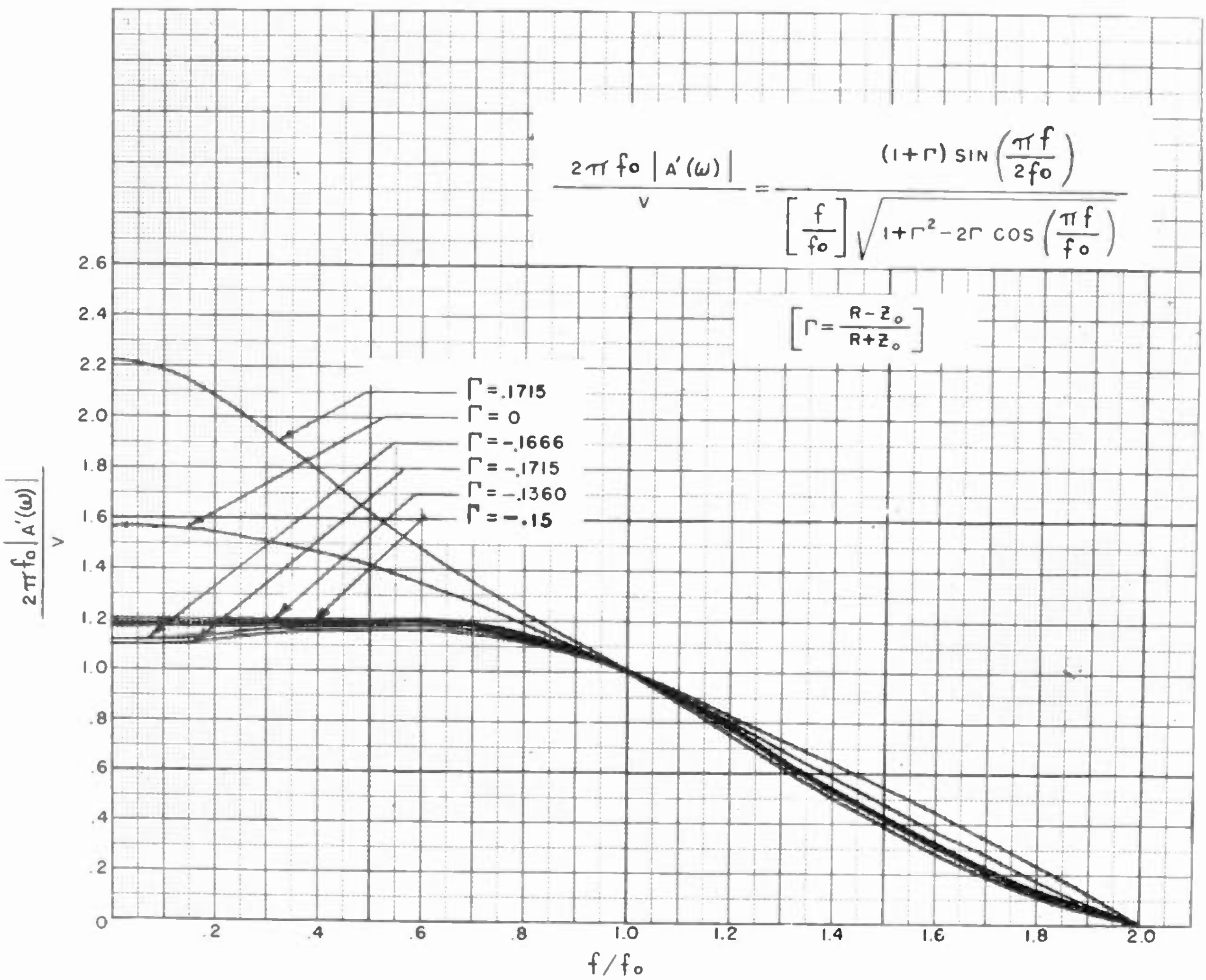


Fig. 2 - Frequency characteristics of idealized form of noise generator.

These can be readily solved by Laplace transform techniques to obtain the spectral and phase equations across R shown below:

The Normalized Spectral Amplitude:

$$|A'(\omega)| = R|A(\omega)| = R|I(0,\omega)| = \frac{V(1+\Gamma)\sin\left(\frac{\pi f}{2 f_0}\right)}{2\pi f_0 \left[f/f_0\right]\sqrt{1+\Gamma^2-2\Gamma\left(\frac{\pi f}{f_0}\right)}}$$

$$\Gamma = \frac{R-Z_0}{R+Z_0}$$

f_0 = RESONANT FREQUENCY OF CHARGED LINE

and the Phase Characteristics:

$$\phi = \pi(1-f/f_0) - \tan^{-1}\left[\frac{\Gamma\sin\left(\frac{\pi f}{f_0}\right)}{1-\Gamma\cos(\pi f/f_0)}\right]$$

For simplicity these equations have been plotted and the results are shown in Figure #2. The curves have been normalized for voltage output load and resonant frequency of the open circuited transmission line. Various values of the reflection coefficient at the load end have been used to show the variation in the output characteristics as a function of varying the ratio of load resistance to characteristic impedance. It can be seen that for a reflection coefficient of -0.15 that a constant spectral distribution is valid to within a few per cent to a frequency in excess of 65% of the cutoff frequency. The phase characteristics have also been plotted for the idealized form of impulse generator and are shown in Figure #3. From a theoretical analysis it appeared readily feasible to develop the type of device required for receiver performance measurement.

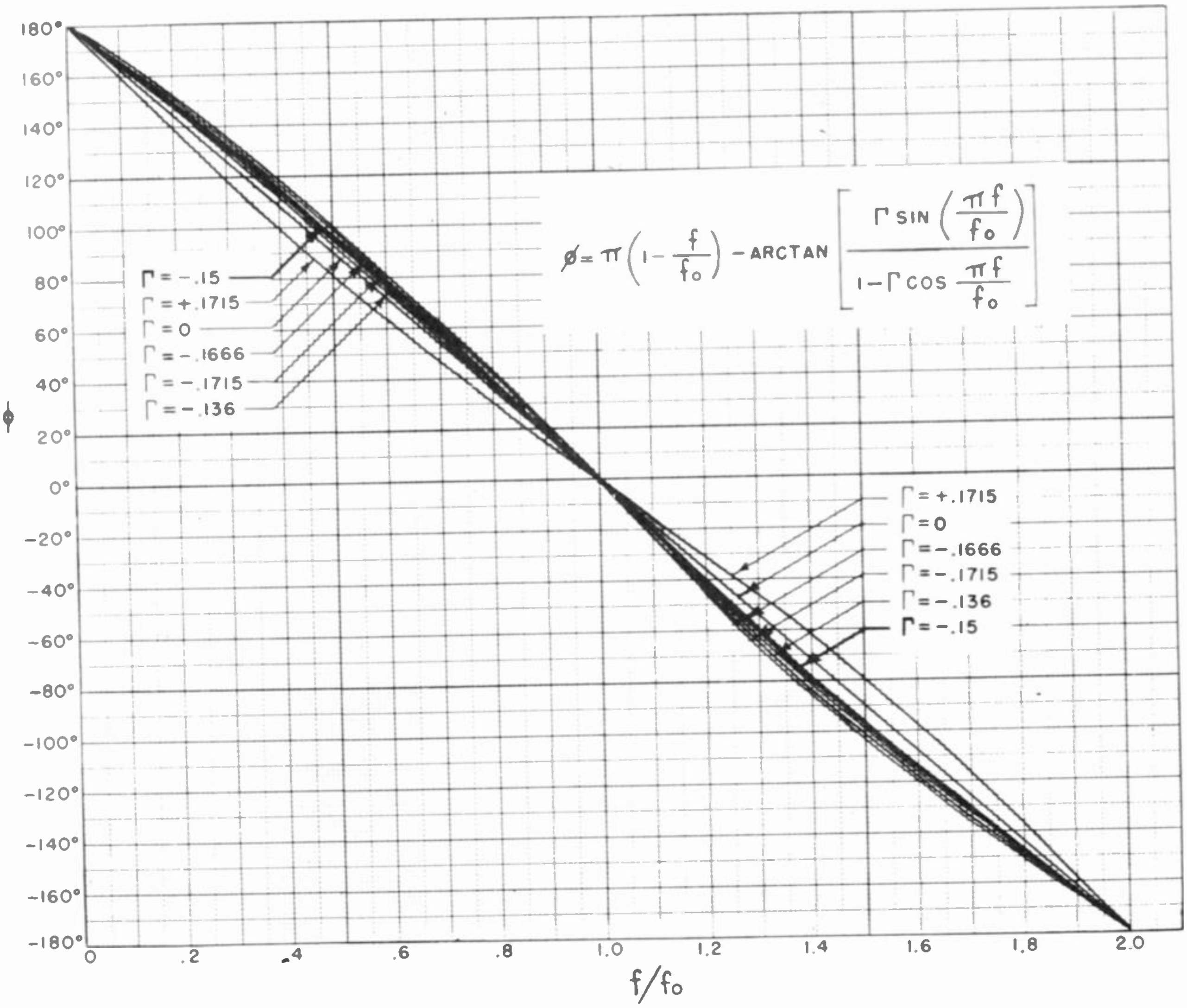


Fig. 3 - Phase characteristics - idealized form of noise generator.

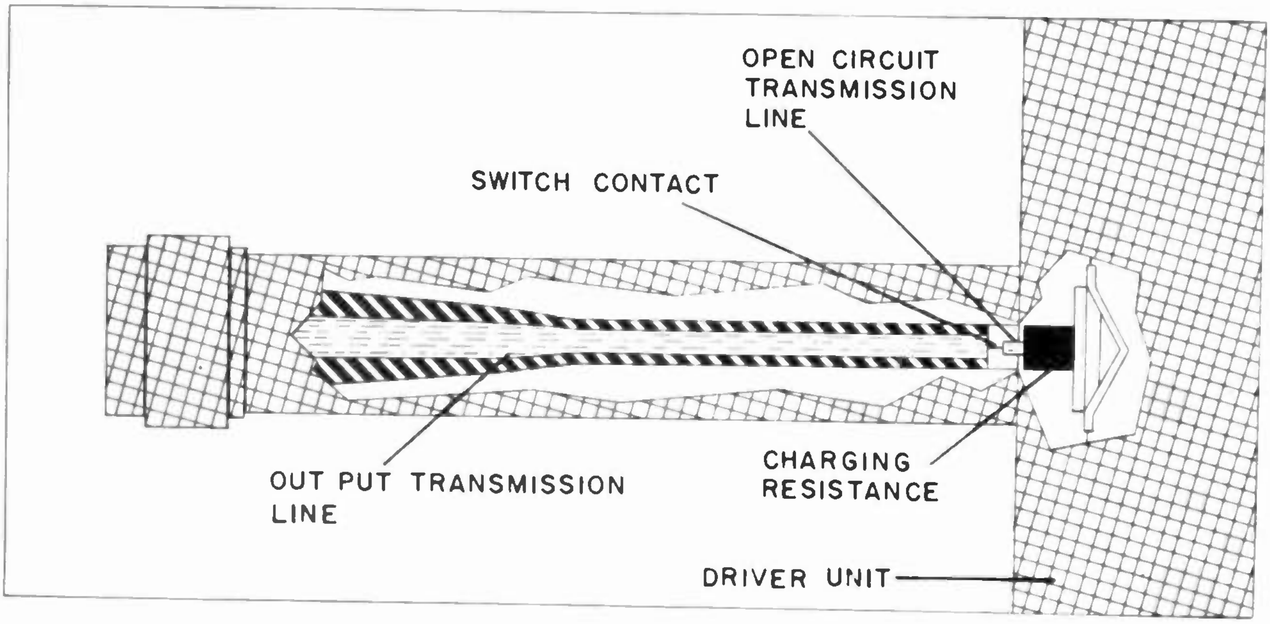


Fig. 4 - Idealized mechanical noise generator.

PHYSICAL CHARACTERISTICS OF IMPULSE GENERATOR

The impulse generator basically consists of a transmission line of characteristic impedance Z_0 which is periodically terminated by connecting it to an output transmission line of characteristic impedance Z_L which is in a matched load (i.e., a receiver thru a directional coupler). In one of the models (Figure 4) the open end transmission line consists of a very small coaxial line designed to be resonant at 35,000 mc/s. The transmission line is charged to the voltage V through a charging resistance of several megohms which is in permanent contact with the inner conductor of the open end of the line. The presence of this resistance across the transmission line has negligible effect on the performance since its magnitude is so large as to appear to the rf propagated wave as an open circuit. It serves the purpose of limiting the charging current to a very small value and at the same time provide a dc charging path which is an rf open circuit.

The output transmission line is tapered very gradually, to a size suitable for coupling to standard fittings. The same technique is equally applicable to radial transmission lines, coaxial transmission lines, parallel plate waveguides, etc., with appropriate physical changes to conform to the field conditions in the particular line used.

DESIGN AND FABRICATION PROBLEMS

The impulse generator would appear to be a fundamental measurement device since the output may be precisely expressed as a function of the charging voltage, the resonant frequency of the transmission line, the associated impedances and the bandwidth of the receiver into which it is fed. While this is the case, any particular noise impulse generator must be calibrated to determine its output characteristics, since this is simpler than measuring the resonant frequency or the impedance precisely. The reliability as a

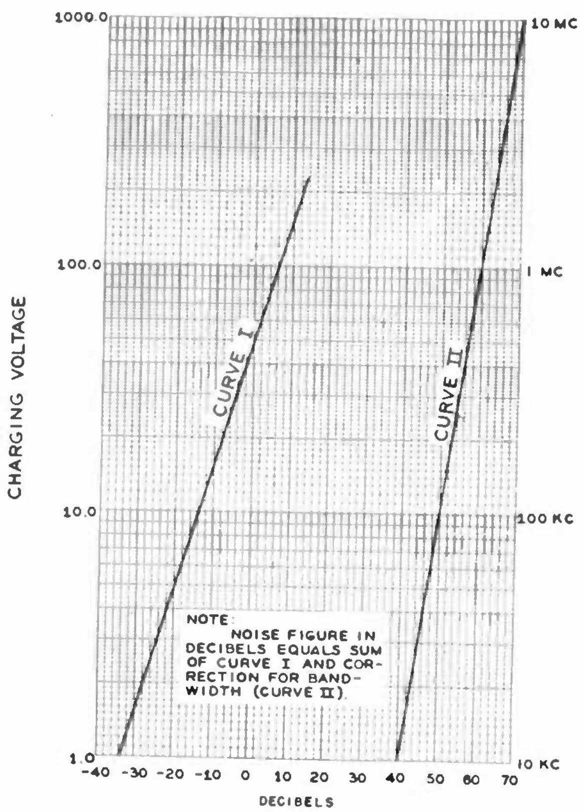


Fig. 5 - Power output vs. voltage.

measuring device is determined in large measure by the stability of the charging resistance and the impedance of the charged line, in addition to the effectiveness of the switching mechanism that terminates the charged transmission line into the output transmission line. A consideration of the coaxial impulse generator will indicate the nature of these problems.

THE CHARGING RESISTANCE

The charging resistance is in the order of several hundred megohms and has been selected so that the time constant of the charging resistance in combination with the capacitance of the charged line is ¼ or 1/5 of the time between switching pulses. This requirement arises from the necessity of keeping this resistance so high as to make it appear as an open circuit to the rf wave. Small physical changes in the size of the contact surface between the center conductor and the charging resistor results in large changes in the potential across the charged line at the instant of discharge. This produces at the output of the receiver a pulse, whose amplitude varies and makes measurement difficult, if not impossible. The power output and the measured noise figure varies as the logarithm of the voltage change which varies inversely with the charging resistance. The magnitude of the power output as a function of the change in voltage is shown in Figure 5. Great care is required in the assembly to eliminate this source of trouble.

CHARACTERISTIC IMPEDANCE OF THE CHARGED LINE

Since the ratio of the impedance to the output line determines the nature of the decaying pulses and is the criteria for determining the output energy and its spectral distribution, a satisfactory impulse generator requires that this ratio be held constant. If the supporting member for the charging resistance and charged line is not adequately designed, the eccentricity of the center conductor with respect to the outer conductor will vary from triggering pulse to triggering pulse. Since the dimensions involved are in the order of thousandth's of an inch, a change of 0.001 inches while in itself small, will produce a large change in the characteristic impedance of the transmission line. The resultant output will then vary erratically because of the variation in position of the inner conductor caused by the

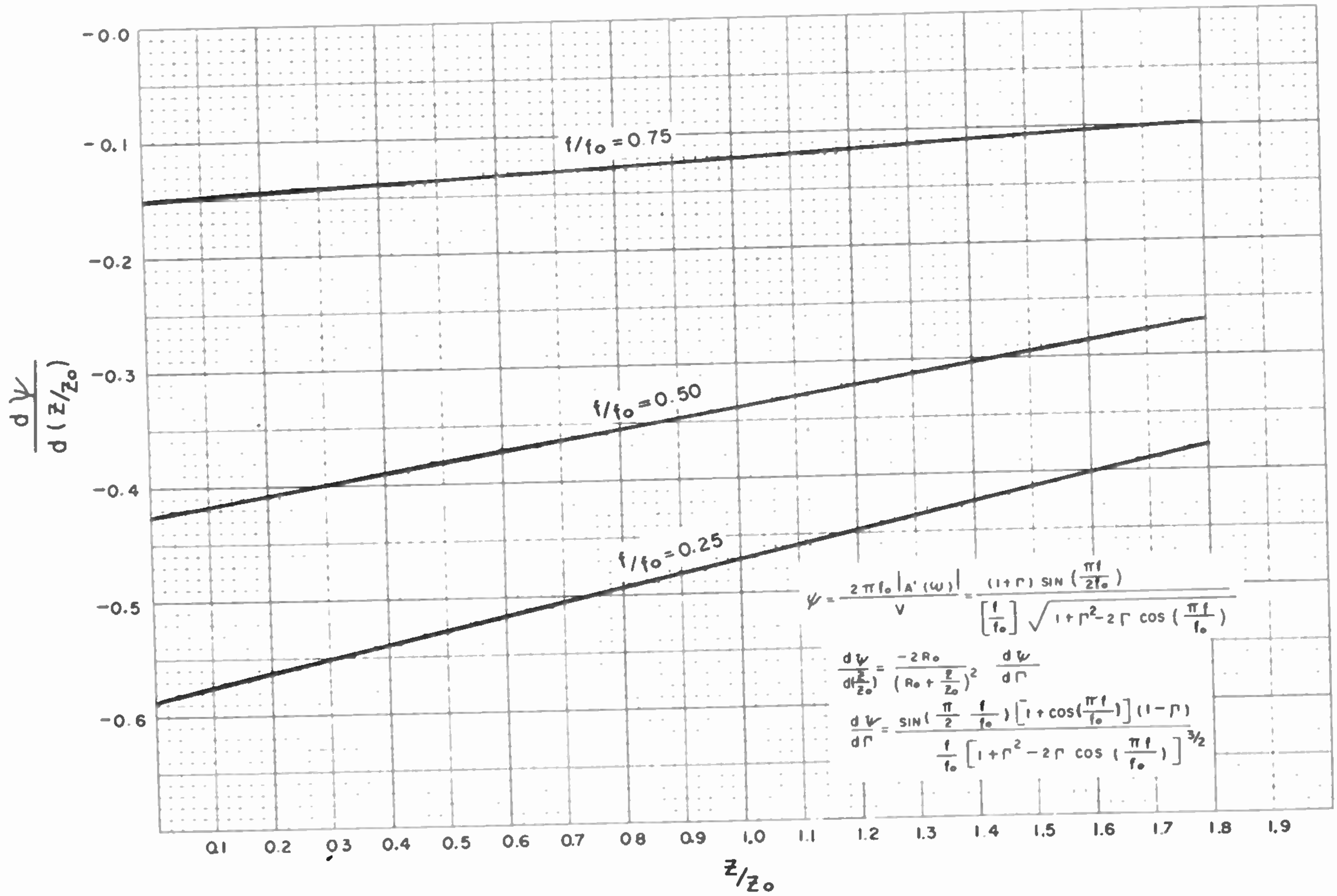


Fig. 6
Spectral amplitude variation as a function of changes in the characteristic impedance of the charged transmission line.

driver mechanism. This problem should receive considerable attention from anyone attempting to fabricate impulse type generators. The change in spectral amplitude as a function of the variation in the impedance of the charged line is shown in Figure 6.

CONTACT RESISTANCE

The contact resistance manifests itself as a change in the apparent output impedance. Since the contact resistance is only a fraction of a percent of the load impedance, the change in output resulting from nominal variations in contact resistance will not be serious, as may be seen in Figure 7. However, the closing of the contact will produce time jitter in the received output pulse if the switching mechanism is not precisely adjusted to make contact in time synchronism with the driving trigger.

FABRICATION CRITERIA

The critical items requiring special attention on the part of those attempting to fabricate impulse generators are the following:

a. Stability of charging resistance
b. Concemtricity of center conductor of charged line.
c. Timing precision and goodness of switching contact.

APPLICATION AND MEASUREMENT TECHNIQUES

The impulse generator is most useful in the continuous monitoring of pulse receivers where the optimum performance is essential and the receiver can not be disabled from its normal functioning to check its performance. The impulse generator can be used during the dead time of the reciever under normal operating conditions since the power output of the generator is sufficient to permit measurement thru a directional coupler. The independence of frequency has proved particularly useful in the case of tunable or multiple radar receivers. Two measurement techniques have been recommended for radar applications. The impulse generator charging potential is adjusted to give an output 10 db above the design noise figure of the receiver at the receiver input terminals (i.e. 10 db plus the attenuation of the directional coupler and cables above the design noise figure). A switching trigger is furnished so as to produce the impulse at a suitable time.

The performance can then be measured at the receiver output by either of the following meth-

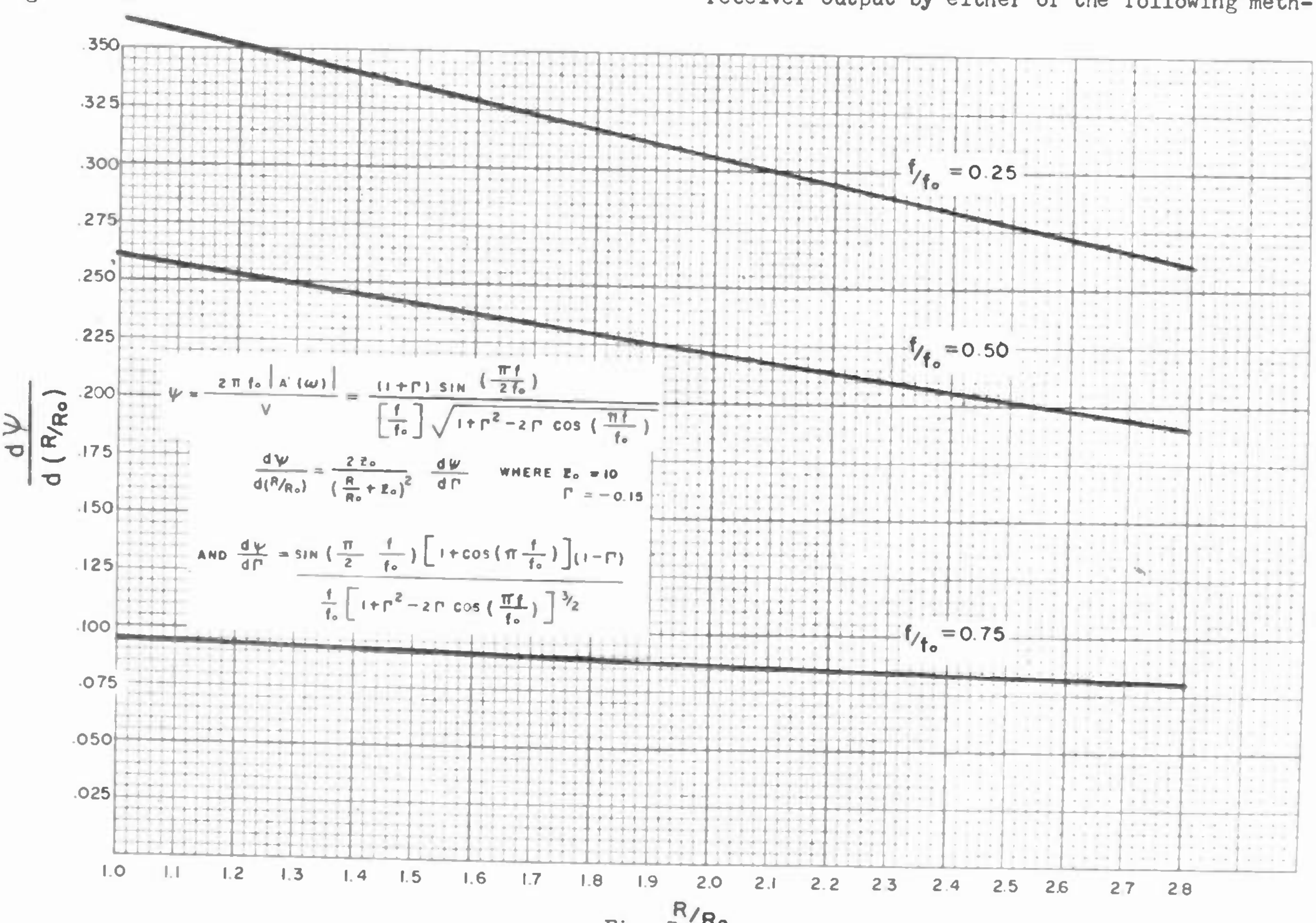


Fig. 7

Spectral amplitude variation as a function of changes in the contact resistance (normalized to output impedance R_o).

ods: An oscilloscope is calibrated in decibels above a reference level. The gain of the oscilloscope is adjusted to bring the peak of the impulse indication to the 10 db point. The deviation of the noise figure from the design value can then be read directly in decibels by observing the location of the average noise peaks. Alternately, a gating circuit and logarithmic meter or recorder can be used to alternately measure the peak of the impulse and the average noise peaks. The ratio in decibels subtracted from 10 is the deviation of the receiver from design performance. In military applications a limit is set on the allowable deviation before maintenance is mandatory.

EFFECT OF BANDPASS PHASE CHARACTERISTICS

The effect on the output amplitude for a unit impulse as a function of the bandpass amplitude characteristics have been investigated analytically for rectangular, sinusoidal and Gausian bandpasses as typical of idealized models of receiver types. These results have been plotted in Figure 8 for the case of identical bandwidths as normally defined, 6 db in voltage. Shown for comparison in Figure 8 is the response for a unit pulse of width $\tau = \frac{2}{\Delta f}$

The equations for the amplitude response for an impulse thru a rectangular bandpass with arbitrary, linear and discontinuous phase characteristics are given in Table I. With arbitrary phase response the output signal consists of the product of "echoes" that are delayed, and a dispersion of the time solution results. With linear phase response $\phi(\omega)=K\omega$ the solution shows that the original signal is merely delayed and modified by the amplitude characteristic but no distortion exists. For a phase discontinuity at w_0 the peak signal amplitude decreases as a function of the phase angle of the discontinuity. The latter two cases are plotted in Figure 9. The effect of phase characteristics on the impulse response has been shown to correspond to that experienced by a rectangular pulse passing thru the same receiver. Since this effect is not present when random noise is used to measure receiver performance it would appear desirable to use an impulse generator for performance measurement of receivers subject to phase changes such as stagger tuned i.f. strips.

TABLE I

(a) FOR ARBITRARY PHASE CHARACTERISTIC $\phi(\omega)$

$$v(t) = \frac{2A(\omega_0)}{\pi} \prod_{n=0,\,k=-\infty}^{n=\infty,\,k=+\infty} J_k(b_n)\cos\left[\omega_0\left(t+\frac{2\pi n}{\Delta\omega}+\theta\right)\right]\left[\frac{\sin\frac{\Delta\omega}{2}\left(t+\frac{2\pi n}{\Delta\omega}\right)}{t+\frac{2\pi n}{\Delta\omega}}\right]$$

$$\omega_1<\omega<\omega_2 \qquad \Delta\omega=\omega_2-\omega_1$$

$$\phi(\omega)=\sum_0^\infty b_n \sin\frac{2\pi n\omega}{\Delta\omega}$$

(b) FOR LINEAR PHASE $\phi(\omega)=K\omega$

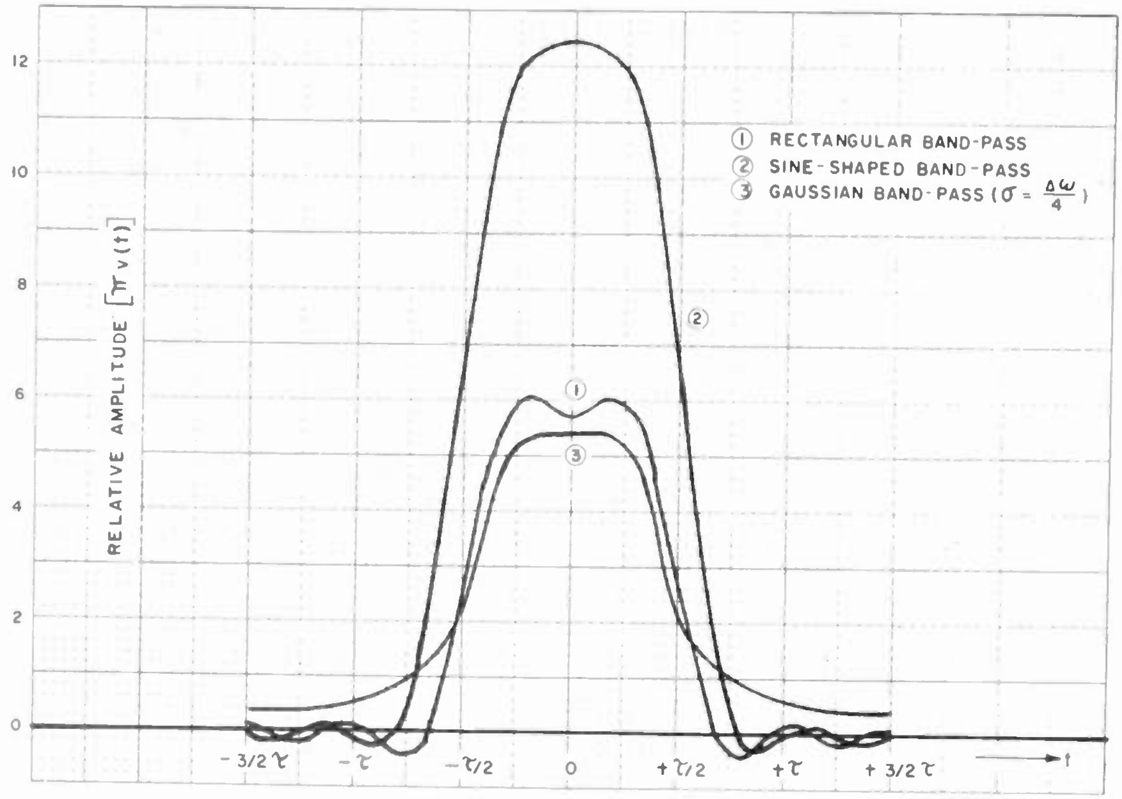
$$v(t) = \frac{\Delta\omega A(\omega_0)}{\pi}\cos\left[\omega_0(K+t)+\theta\right]\frac{\sin\frac{\Delta\omega}{2}(K+t)}{\frac{\Delta\omega}{2}(K+t)}$$

(c) FOR PHASE DISCONTINUITY AT ω_0

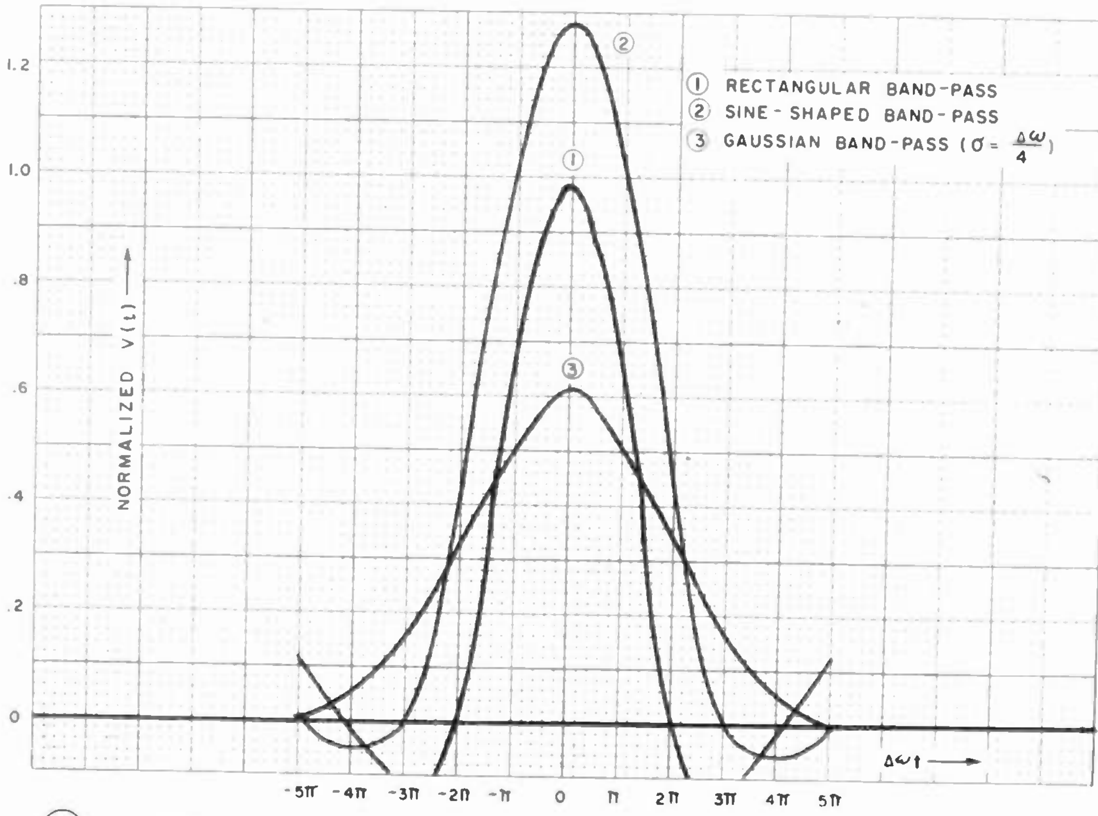
$$\phi(\omega)=-\theta \qquad \omega<\omega_0$$

$$\phi(\omega)=+\theta \qquad \omega>\omega_0$$

$$v(t) = \frac{\Delta\omega A(\omega_0)}{4\pi}\cos(\omega_0 t+\theta)\frac{\sin^2\frac{\Delta\omega t}{4}}{\frac{\Delta\omega t}{4}}$$



Ⓐ RESPONSE OF RECTANGULAR, SINE AND GAUSSIAN SHAPED BAND-PASS TO A UNIT PULSE INPUT FOR $f = 2/\tau$



Ⓑ RESPONSE OF RECTANGULAR, SINE AND GAUSSIAN DISTRIBUTION-SHAPED BAND-PASS TO UNIT IMPULSE FUNCTION

Fig. 8 - Effect of band-pass characteristics on measured performance.

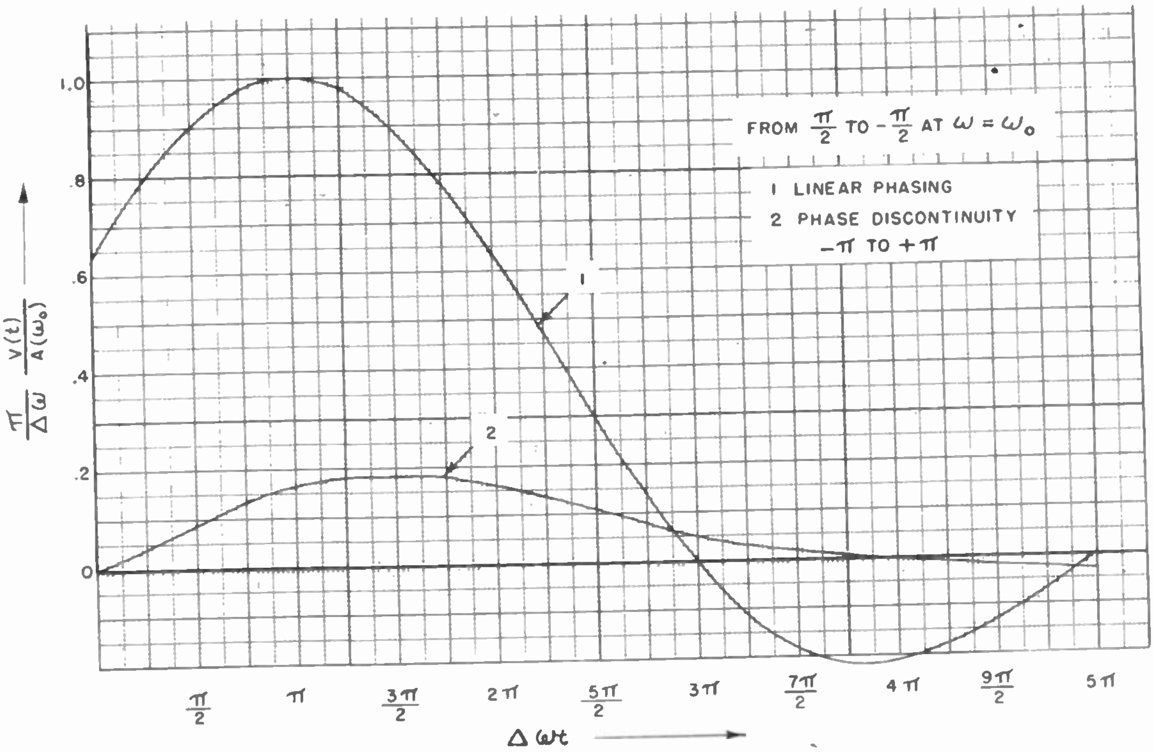


Fig. 9
Output response to impulse generator of rectangular band-pass with linear and discontinuous phase characteristics.

AERIAL METHODS IN MICROWAVE SURVEY

Marc Sheldon
Microwave Services, Inc.
New York, New York

Lewis A. Dickerson
Lockwood, Kessler & Bartlett, Inc.
Great Neck, New York

Summary

The reasons for the importance of aerial survey methods in the design of microwave systems are explained, and a brief history of aerial survey is given as background material.

The objectives of a microwave survey are considered in terms of the nature of the information required. Five specific types of aerial survey are described in detail, including the equipment used, procedures followed, nature and accuracies of results obtained, and limitations imposed by each method. Aerial reconnaissance (visual), aerial reconnaissance (photographic), radar profiling, aerial photogrammetry, and path testing using helicopters are described.

Under the "Applications" section the foregoing methods are discussed in terms of the use to which each is suited, and a comparison of these methods with each other and with ground survey techniques is made.

Introduction

The application of aerial methods to microwave survey work is still in the early stages of development, (partly because the microwave art itself is still fairly new, particularly in its commercial applications.) This discussion of aerial methods in microwave survey is prompted by the increasing popularity of their use, and the need for detailed information on what the various methods have to offer, the factors to be considered in determining which method to use in preference to or in conjunction with other methods, and the results to be expected. The procedures used in ground surveying are considerably older and better known than aerial survey procedures, and for this reason ground survey procedures are not covered except as required for the sake of comparisons.

The history of aerial survey begins with aerial photography from kites and captive balloons in the 19th century, and continues in the early 20th century, with visual and photographic reconnaissance from airplanes. This gained important impetus from World War I when aerial reconnaissance proved its strategic and tactical value in military operations.

The art progressed further in the 1930s in the United States when instruments were added to the aerial cameras, so that it became possible to obtain more precision type data from photographs which had previously been treated as maps as they stood.

During World War II applications of the instrument methods in connection with aerial survey experienced a tremendous expansion and American manufactured equipment became available for the first time. Immediately after the end of World War II considerable refinement and development took place as many of the war time electronic developments were coupled with the instruments used in aerial plotting.

Objectives of Microwave Survey

In locating stations in a microwave relay system, the general objectives are as follows:

1- To specify the minimum number of relay stations consistent with the designated points of communication and satisfactory overall operation.

2- To select sites which involve minimum expense for:

a- Rental or purchase of the land
b- Site improvement (e.g. drainage)
c- Access Road construction
d- Commercial power installation
e- Towers (minimum height)
and which have
f- Position of minimum weather hazard (e.g. extremes of wind or ice)
g- Location favorable for other considerations of the communications system, e.g.VHF coverage.

Obviously, these optimum conditions will rarely occur conjointly at particular sites. It is therefore necessary to strike the best balance among them in locating relay stations.

(A common example of the problem is the site which is ideally located from the point of view of elevation, but which requires costly access road and commercial power to be brought to it.)

Microwave Path Requirements

Let us consider the requirements of a microwave path, a single hop between two repeaters.

1- Clearance - Satisfactory microwave operation requires that there exist line of sight between the two antennas or passive reflectors. In addition to this, some fraction of the 1st Fresnel Zone radius as a minimum clearance of the line of sight path above any obstruction point is required. The clearance requirement must take into account all obstructions, including trees, buildings, and any other natural or man-made objects.

2- Favorable propagation conditions- Such terrain features as bodies of water and smooth surfaces may prove to be a source of reflections, which under certain conditions may cause serious fading and distortion of the signal.

Bodies of water may also produce atmospheric conditions conducive to tropospheric ducting, a phenomenon of non-linear refraction.

3- The path length must not exceed the distance over which the attenuation can be made up by equipment and antenna gain. This is rarely the determining factor in limiting the length of the hop; practical tower heights to obtain clearance will generally come into play first, except where long paths between high ground elevations can be realized without high towers.

Survey Information

The known factors in laying out the microwave system are generally the following:

1- Designated points of communication, including remote metering and control points.

2- Nature and extent of auxiliary communications, e.g. VHF mobile radio.

The facts which the survey should bring out include:

1- The elevations and positions of terrain features between fixed points of communication, covering an area band sufficiently wide to include all locations which may be considered for sites, and all points which may constitute obstructions on the final path.

2- The nature of the terrain in this band, including information concerning tree heights, bodies of water, character of earth surface and similar factors which will affect both clearance and propagation conditions.

3- Man made features, such as roads, buildings, railroads, power lines. This type of information is more important in the vicinity of proposed sites than on intermediate portions of the path (except insofar as it influences clearance considerations). In the immediate area of a proposed site, detailed data concerning subsurface conditions for access roads, land availability, drainage, etc. should be gathered.

Up to now, nothing has been said of the specific accuracies to which the information required is needed. This is important, because it affects the way in which the survey is conducted.

To begin with, the accuracy needed is not the same throughout. In determining elevations in the area band between fixed points, a preliminary survey of a relatively low order of accuracy will often suffice to indicate the likely locations for the repeater stations. Once these are determined, more accurate measurement of the elevations in the several proposed paths will be made to determine the best conditions or to confirm tentative findings. At the same time, the other detailed information concerning the path and the sites will be gathered.

This procedure of working from rough data to more precise data is very often followed, and is the basis for the approach to the survey problem.

Aerial Methods - General

The determination of when the use of aerial methods in microwave survey is advantageous, and which methods are to be preferred under which conditions, will become fairly clear as the methods are discussed in detail. However, the following general statements will apply in most cases:

1- In terrain where travel is difficult, aerial methods offer obvious advantages over ground methods of survey, including that of speed.

2- Some ground work is required in

conjunction with all types of aerial survey, the kind and amount depending on the particular aerial method used and the terrain in question.

Aerial Reconnaissance - Visual

To obtain a general impression of the nature of the terrain to be surveyed, regardless of the eventual method of survey to be used, a visual aerial reconnaissance can be very valuable, and can generally be undertaken at minimum cost.

The route is generally flown at low altitudes, and spot photographs may be taken of significant terrain features. This is especially valuable if the area under consideration is scantily mapped; it may help determine what type of final survey should be made.

Aerial reconnaissance is valuable for more than terrain information. Such important data as the road network of an area, power facilities, and the like can be observed directly and by inference. The general condition of the land and the use to which it is being put may provide valuable information as to the ease and cost of securing a site. For example, a site in an uncultivated, sparsely populated section could probably be obtained at a much lower figure than a similar site in a more populous urban area, or in rich farmland.

The object of the reconnaissance is to gather as much information as possible which will aid in narrowing the choice of potential station sites. Of the necessary information the most difficult to obtain from this type of reconnaissance is terrain elevation data.

One method which has been used to select probable paths is to take the airplane up to a point where the horizon is about thirty miles,and head it in the general direction of the required path. By careful sighting on terrain features in the distance, and using experience as a guide to the nearer and farther points, (distant hills take on different colors from closer ones) likely hills or other high points may be noted. The plane can then fly to such points to identify them, and these can later be checked as potential sites by one of the methods to be described.

Aerial Photographic Reconnaissance

Although aerial photographic reconnaissance is a reconnaissance rather than a final type of survey operation, it provides a wealth of data which will aid the engineer in reducing the number of suitable paths for each microwave hop to a relatively few, which can then be finally checked by other means. This is secured by covering the area of interest with vertical aerial photographs, a stereoscopic study of these photographs, and a correlation of this study with all other existing information concerning the area.

In the field the work involved in the aerial photographic method consists of covering in a systematic manner with vertical aerial photographs the entire area of interest. Although oblique aerial photographs may be of some value in a few special cases, their use is quite similar to visual aerial reconnaissance and they are not considered further here. For vertical photography an aerial camera is mounted in an airplane with the axis of the camera pointing vertically downward through an opening provided in the floor of the plane. Figure 1 shows a view of such a camera mounted in an airplane. The mounting provides for small adjustments in the position of the camera to maintain it as closely as possible in the vertical position and provides for large rotations around its optical axis to adjust for the drift angle of the plane. This is the angle between the actual heading of the aircraft and the resulting ground track and is the result of cross winds. By reference to level bubbles mounted on the camera, the operator is able to keep the camera axis vertical within about one degree. With tilts of this small magnitude the resulting photographs closely approximate planimetric maps of the area covered. As the plane flies along a line as straight as possible and at as constant an altitude as possible, the camera operator makes successive exposures. The time between exposures is so arranged that each exposure overlaps the area of the preceding one by 55 to 60%. If more than one strip, or flight, of photographs is required to cover the area of interest, parallel strips are flown and so placed that the area covered by one side laps the adjacent one by 20 to 30%.

There is a number of different types of aerial cameras suited to reconnaissance work. As the purpose is usually to reconnoitre as large an area as possible with a minimum of flying effort, it is desirable that the angular field of coverage of the camera be as large as possible. At the same time, it is desirable that the focal length of the camera lens be great enough so that the resulting images are of a usable size. The optimum combination at present is provided by a camera having a focal length of six inches and exposing a

negative nine inches square. This camera covers an angular field of about 93° across the diagonal and about 74° across the width of the negative.

Altitude and Scale

In choosing the altitudes for flying reconnaissance photography a compromise of two desirable objectives is involved. It is desirable that the altitude be as high as possible as each photograph thus covers more area and economy in flying and later use of the photographs is thus achieved. Conflicting with the high altitude, however, is the scale of the resulting photographs. As altitude goes up, the scale of the photographs goes down in a linear proportion. A scale large enough must be retained to secure the desired information from the photographs. On the average, a flying height of 20,000 feet above the ground will be found optimum. At this height the scale of the photograph is 1:40,000 or about 3,300 feet to the inch. A single flight at this height covers a strip of ground about six miles wide and photographs are exposed at intervals of about two and a half miles along the flight.

Weather Effects

Brief reference should be made here to the effects of weather on securing the required photographs. They are only flown at times that clouds are completely absent, when the ground is not covered with snow and when the sun is at an altitude of about 30° or more. In the northern portion of the United States this results in some two or three months of the winter not being usable. The occurrence of cloudless conditions varies with season and location but is not as great a hindrance as it might appear. On the average we can expect suitable conditions to occur about every two weeks and in the worst areas we should expect a suitable day about once a month. One such day will usually be sufficient to cover an area for microwave purposes.

Use of Photographs

After processing of the aerial negatives and the preparation of contact prints on paper, the remainder of the reconnaissance work in this method takes place in the office. The overlapping photographs are matched one to another and stapled on to large panels of wall board or similar material. They may be retained in this form for work purposes but it is usually more convenient to copy the boards down to about one fifth size. Prints made from such copies are known as photo indexes and actually constitute a composite picture, or mosaic, of the area of interest having the properties of a rough planimetric map. Figure 2 shows a reduced portion of such a photo index. The individual aerial photographs are then taken off the boards and used singly or in small groups for close study of the areas covered. The photo index serves as a rough map and an index of the relative positions of the individual photographs. It also provides a good base upon which to record the critical items developed in the reconnaissance.

Use of Other Data

To be of most value, the aerial photographs and indexes are used in conjunction with all other data that can be assembled for the area being studied. These data include all possible maps of the area, information on the existing survey control in the area, location of power lines, etc. With such data at hand, an engineer experienced in topographic mapping from aerial photographs can prepare a fairly complete study of the area in question to serve the needs of the engineer designing the microwave system. The work is best done with these two engineers working together. As indicated by the name of the method, the resulting study is only a reconnaissance. THe final information needed must be secured by other methods although this reconnaissance reduces very greatly the overall work required.

Information Obtained

It has been indicated that this reconnaissance results in a fairly complete study from the standpoint of microwave work. Let's examine the type of data we can secure and how. Figure 3 is an enlargement of a portion of an aerial photograph covering an area in Wisconsin and flown at a height of 30,000 feet. It has been enlarged here to demonstrate how much information is contained in the original and to offset the loss of detail due to reproduction here in half tones. (The 20,000' altitude photography previously indicated as optimum would provide even more detail, although the strip covered would be narrower.) The first important item should be fairly evident to you. This is the road network. The photographs show clearly the roads in the area and in rough map form. Thus we can readily determine how accessible a proposed tower site is for construction and maintenance. Stereoscopic study of overlapping pairs of photographs also enables us to determine feasible routes across country into sites where they are

desirable away from roads. Cultural development, buildings etc., are also quickly apparent on the photographs.

With power line data at hand, it is usually possible to locate the position of such lines on the photographs and thus determine the availability of power. In some parts of the country it is possible to determine where property lines are likely to be located on the photographs and an estimate may be made of the likelihood of securing rights-of-way and easements. The photographs show clearly the existence and location of forested areas and bodies of water and their effects on microwave installations may be readily evaluated.

Probably the most important information required is that concerning the elevations of the terrain and objects on the terrain and it is here that the photographic reconnaissance is weakest. But even here the experienced topographic engineer can provide a wealth of information, some of qualitative value only but with approximate quantities in other cases. In a few cases, this information may result in a definite indication of the most feasible horizontal location for the microwave path. Only in an isolated case can it be expected to furnish information to determine final design data. Information from other methods will almost always be necessary for this. But in almost all cases the method will indicate the likely horizontal locations for microwave paths and permit rating these in order of feasibility.

Correlation of Data

The key to securing elevation information by the photographic reconnaissance method is an orderly correlation of all existing elevation data with the topographic features appearing on the photographs. To develop these topographic features completely, stereoscopic viewing and study of the aerial photographs is essential. With the three dimensional view it is possible to identify the high and low ground, the trees that are high and those that are low, the buildings that are of ordinary height and those of extraordinary height, etc. This may be compared with the viewing of 3-D movies or the results of amateur stereo photography. The view obtained from the vertical aerial photographs is similar except that heights are greatly exaggerated since the two overlapping views are taken some two and one half miles apart. One is able to recognize the heights of features amounting to as little as five feet.

So from the photographs it is possible quite easily to outline the drainage network-which is in the low ground-and the drainage divides which are on the high ground. It can easily be seen where the pronounced high points are located that are of concern, and where the high structures and vegetation are located. The greatest difficulty arises in estimating quantitatively the elevations of the features seen.

Accuracies

There is no fixed guide that can be given to indicate what accuracies may be achieved in this. By use of a very simple measuring attachment to the small stereoscope it is possible to do some measuring with relatively good accuracy for features such as cliffs, buildings, or tanks. Figure 4 shows an engineer using one of the simple types of measuring stereoscopes. With this, heights of features such as cliffs, etc. can be determined to an accuracy of about four or five feet. However, in microwave work the concern is more likely that of the terrain elevation at some point with respect to terrain elevations some ten or more miles away. In determining elevation differences over such great distances no simple measuring device working on flat paper prints can be of much value. Large sources of error of unknown and variable magnitudes and directions are present due principally to changes in flight altitude, tilt of the aerial camera and dimensional changes in the paper prints. Elevation differences must be determined by estimation guided by such elevation data as is known for the area and such factors as usual drainage gradients, railroad gradients, etc. As indicated before, no general accuracy can be assigned to indicate what can be expected in this estimation process. Given a specific case, however, the experienced topographic engineer can assign a probable accuracy to his estimates in that case. The accuracies may range from 20 feet to 1000 feet or more.

Photogrammetric Survey

Photogrammetry, which literally means photo-measurement, is essentially the same method which is used in preparation of topographic maps from aerial photographs, the way in which the great majority of topographic mapping is performed today. The principal differences between the application of this method for mapping and for microwave purposes are in the matters of accuracy and in the completeness of the mapping.

The photogrammetric method is based on the principle of taking two photographic views of an object from two different points. When these two are viewed and used together stereoscopically we obtain three dimensional information. Unlike the photographic reconnaissance method where stereoscopic viewing may be applied for certain results, photogrammetry is a precision survey method in which unknown points are determined with reference to known control points, as in ground survey. The survey makes use of the information contained in the images recorded by the camera to determine the positions and elevations of any desired features to a relatively high degree of accuracy.

This method would not be used by itself, but would only follow after a reconnaissance had been performed either on the ground or from the air, visually or by the use of aerial photographs. Since aerial photographs are involved in the method, it is usual that these would be used for at least part of the reconnaissance. The method has three principal phases which are described below.

Photography Phase

The first principal phase of the photogrammetric survey method is that of the aerial photography. This takes place in the field in about the same manner as was described previously for the photographic reconnaissance method. Two significant points of difference should be noted, however. The first concerns the area to be photographed. In addition to the area to be surveyed, the photographic flights must be arranged with reference to the existing ground survey control. The work must be anchored to such control and it must be included in the area photographed in the most advantageous manner. The second point concerns the aerial camera used. This must be a precise mapping camera matched to the photogrammetric equipment to be used. In this country today this generally means that it will have a focal length of six inches and expose a nine inch square negative. As in the case of the photographic reconnaissance method, a flying height of about 20,000 feet will usually be optimum.

Control

The second phase is the procurement of ground survey control to be used with the aerial photography. In all but a few isolated areas in this country this consists of assembling the information on existing control in the area. For the most part, this will consist of horizontal and vertical control established by the U.S. Coast and Geodetic Survey and the U.S. Geological Survey. In addition, survey data for highways, transmission lines and other similar developments may be of value. In only a few isolated areas will it be found necessary to perform new ground survey work in the field, and even here it will consist only of levelling. On the average it will be found in this country that level lines exist at a spacing of 20 to 30 miles. Only rarely will the spacing exceed 50 miles. In these rare cases, a low order of spirit levelling is necessary to provide a spacing not exceeding 50 miles. Horizontal control will almost always be found at spacings under 100 miles, and this is ample.

To be useful in the photogrammetric survey the ground control must be identifiable on the aerial photographs. In all but a few rare cases this identification can be performed in the office from the descriptive data available. Not all points will be identifiable, but those that are will be at intervals sufficiently close to provide the necessary control for the work. Where level lines are in existence it will be found that bench marks were established at about one mile intervals along the lines. As the strip of photographs covers a width in the order of six miles, more than enough bench marks will lie within such a width where a level line crosses it. Sometimes only one bench mark identified in such a width will suffice, although two or three are more usually desired. No difficulty is encountered in identifying points at an interval sufficient to accomplish this. In a rare case it may be necessary to go into the field, recover the bench marks there and identify their location on the photographs. In those rare cases where new levels are performed in the field, the identification of elevation points is made on the photographs at the time they are established.

The same situation is encountered with the identification of horizontal control on the aerial photographs; a sufficient number can be identified in the office from existing data and only in a rare case is it necessary to identify a point in the field. Should field control work be necessary for either horizontal or vertical control, it would be combined with another phase requiring field work so as to minimize the overall work in the field.

Survey Phase

The last principal phase is the actual photogrammetric survey, which

takes place in the office. As noted previously this is essentially a topographic mapping operation. The reconnaissance that preceded it would have resulted in a selection of the features required to be mapped. For example, the reconnaissance may have selected several possible microwave paths and it is next desired to map all features within narrow bands centered on these paths so that profiles may be plotted. In another case the reconnaissance may have resulted in a need for the determination of location and elevation for all prominent heights in a given area. Whatever the case, the mapping procedure is limited to the features selected so as to secure the required information at the minimum cost.

Basic Principles

To describe the photogrammetric method, it is necessary first to go briefly into the fundamental principles. The first of these concerns the aerial camera used to expose the photographs. The precision aerial mapping camera is really a surveying instrument. From its position in space at the instant of exposure, the camera lens actually observes the angles in space among the infinite number of objects in its field of view. Through the lens and film action these angles are recorded on the plane of the negative by the record of the image positions on that plane. The camera is so calibrated that all of its metrical characteristics are determined and from these and the image positions in the negative plane the angles between the various objects can be measured. Cameras are so constructed and calibrated that an accuracy in the order of fifteen seconds of arc is secured in measuring the spatial angles between objects. So in addition to producing photographs that have a pictorial value the camera also observes and records space angles from the successive points in space that the lens occupies along a flight line when the shutter is instantaneously snapped open and closed.

To illustrate the principle underlying the photogrammetric method of using the angle records made by the aerial camera, visualize a fictitious condition. The illustration shown in Figure 5 will assist in visualizing this condition. Instead of one camera flying through space, assume two identical cameras suspended and fixed in space at two successive exposure stations. Further assume that the shutters are so interconnected that they may be tripped with a single impulse.

Examine what happens when the shutters are open. From some given object point in the common overlapping fields of view two light rays originate. One ray passes through each camera lens and creates on each photographic emulsion a latent image of the original object. This same thing occurs for every one of the infinite number of objects in the common fields of view. Now without disturbing the cameras let us imagine that the two negatives are developed, that a source of illumination is placed over each one and that the two shutters are fixed opened. Under this condition a light ray will originate from each of two common image points. These two rays will pass through their respective lenses on into space and will intersect at the exact point occupied by the original object that was imaged. Similarly, this occurs for every pair of common images in the overlapping projections and the composite of all the ray intersections forms a surface exactly the same as that of the original area photographed.

It is necessary now only to reduce this hypothetical situation, which is thousands of feet in size, down to a reasonable working space and to introduce a system of measuring where the ray intersections occur. When this is done we have established the principle of photogrammetric surveying instruments.

To reduce the previously assumed situation to a workable size we have only to move the two cameras, now operating as projectors, closer together. When one projector is moved parallel to itself with the lens moving along the line joining the two lenses, all rays continue to intersect in space at their true relative positions but at a reduced scale. In photogrammetric instruments the spacing between projectors is in the order of one or two feet and the resulting scale may vary from about 1:500 to about 1:30,000, depending upon the flying height.

Measurement Procedure

Measuring systems are of several types but all provide for placing a measuring mark on the intersection of the rays in space and measuring its position along three mutually perpendicular axes. These measurements at the known scale at which the intersections are occurring result in the horizontal positions and elevations of the objects on the ground surface. Illustrated in Figure 6 is one of the simpler photogrammetric instruments. Note that it has two projectors which are closely matched replicas of the aerial camera. They are so mounted that they can duplicate every motion that occurred to the aerial camera in flight. They

project their cones of rays down into space and the corresponding rays intersect in space some six inches above the table surface.

The measuring system is made up of the table surface itself, which is actually a surface plate, and the small movable stand resting on the table. In the center of the white disc on that stand is a measuring mark which is brought into coincidence with ray intersections by stereoscopically viewing it simultaneously with the photographs projected onto the surface of the disc. A pencil point below the mark transfers the horizontal position to the table surface and a vertical scale on the stand measures the elevation of the mark.

In practice the correct relative setting of the two projectors is determined by observing the ray projections. The projectors are relatively adjusted until all common rays do intersect. This creates the "stereoscopic model", usually referred to as the "model". When this model is created, all effects of airplane and camera tilt, change in flight altitude, change in flight direction, etc. have been compensated for.

To determine the scale of the model and its orientation with respect to the table surface, which is really the datum plane, reference must be made to at least one known horizontal distance and three known elevations in the model. Once adjusted to such known control, the positions and elevations of new points along the edge of the model may then be measured. These then become known control for the next model which overlaps that edge. In this manner, control may be carried through a series of models along a strip until more known ground control is encountered. Accumulated errors are determined at such closing control and give the basis of adjusting the work back through the strip. This carrying of control through a strip is termed "aerial triangulation". Figure 7 shows a more complicated photogrammetric instrument capable of a higher order of precision and of a type most used for aerial triangulation.

Aerial triangulation is the key to the use of the photogrammetric survey method as applied to microwave location. This triangulation is carried out from the ground control that was referred to earlier. As the triangulation is carried out, positions and elevations are determined for all points for which the reconnaissance indicated information was needed. As the triangulation is completed and adjusted the data are immediately at hand to make the final path selection. In some cases more detailed information may be found necessary along all or portions of the path. For this purpose individual stereoscopic models may be reset in the instruments to the previously triangulated control and the detailed measurements are made. With these models complete topographic maps may be prepared, but this will seldom be found necessary except for some limited areas of concern. The resetting of models to the triangulated control is rapidly accomplished and permits obtaining additional measurement data in an expeditious manner.

Accuracy

The accuracy that is achieved by the photogrammetric survey method varies with the amount and spacing of ground control that is used in the work. In all except the very few isolated areas in this country where the existing vertical control is extremely sparse this method will determine elevations correct to within 20 feet as referred to sea level datum. These maximum errors may be expected at points farthest removed from known control. Relative elevation errors over horizontal distances up to about five miles will not exceed 10 feet. Absolute errors in horizontal positions as referred to the national geodetic net may be expected up to about 200 feet. However, relative errors in horizontal positions over distances up to about 25 miles would not exceed about 75 feet and this is the more likely case of concern in microwave work. These vertical and horizontal accuracies are well within reason for microwave work. The accuracies may be improved considerably by performing additional field survey control work but this will generally be found to be uneconomical due to the high cost of field work.

Cost

Costs of the photogrammetric survey method are relatively high. However, it is still an economical method within its sphere of application which is in those unmapped areas where it is necessary to secure survey data on large numbers of points widely spaced throughout the area. In this sphere of application it is really topographic mapping that is required and this is always a costly operation. It is dangerous to indicate costs here since only a very rough estimate figure can be given which will be subject to many variations. In spite of this, it is believed of value to give some indication of order of magnitude. For this purpose the microwave engineer may think in terms of $25.00 per lineal mile of photographic strip involved, of which about $5.00 is

chargeable to the aerial photography. It is pertinent to note that the figure given amounts to only some $5.00 per square mile of area covered, since each lineal mile includes an area 5 to 6 miles wide.

Rate of Work

The "photogrammetric surveyor" with his instrument can carry out his instrument work at a rate of about ten to fifteen linear miles of photographic strip per man-day. In addition to the instrument operator one other man would be required to prepare the necessary preliminary materials and control and to adjust and prepare the results of the instrument work. These two men can begin to produce usable results in about one week after receipt of the photographs and the control data, their working materials. It should be noted that the ten to fifteen linear miles would embrace 60 to 90 square miles. The total rate can of course be stepped up by using more manpower and more instruments.

Availability

The photogrammetric survey method requires the use of expensive equipment, facilities established especially for photogrammetric work and engineering personnel experienced in this specialized field of work. There are large organizations of such personnel with the necessary equipment and facilities in federal government agencies engaged in mapping but these are not available to private firms likely to need such services in connection with microwave installations. There are some ten private firms presently located throughout the country with facilities suited and available for such work. It is an expanding field in which more firms are continually entering, and older firms are enlarging their facilities. No difficulty should be encountered in finding this survey method available to microwave engineering.

Radar Profiling

The basis of this method is the use of radar equipment installed in an airplane to measure the distance from the plane to the ground (or intervening objects, such as buildings).

Since the measured ground elevations must be made independent of the altitude of the plane (which does not remain constant in flight), each reading of the radar altimeter must be correlated with a simultaneous reading of a barometric altimeter, which is itself referenced to a known elevation.

Use of radar profiling must be preceded by some general knowledge of the area, either from existing mapping or from reconnaissance data. This is required in order to select a few tentative paths to be profiled (from the theoretically infinite number of possible lines).

Mapping Conditions

If there is usable (but not necessarily precise) topographical mapping in the area, the preparatory work in laying out the lines to be flown may be based on these maps, used in conjunction with whatever planimetric mapping and other data may be available. For example, the topographical mapping in certain parts of this country dates from 1890 or earlier, and is not sufficiently accurate in either elevation or position data to be relied on for final work. However, it generally gives enough information to indicate which are the probable optimum paths which should be checked by radar profiling. When the road and other information from more recent planimetric maps (of the so-called "county maps" type), and possibly other data from sources such as power line construction maps are added, the selection of a relatively few likely paths can often be readily accomplished.

In an area where there is no usable topographic mapping, a photo-reconnaissance would probably be made as a first step. (In some cases it is possible that a visual reconnaissance alone might be adequate.) From the photographs, together with all descriptive data on the area which can be assembled, much information on elevations, roads and other factors of interest can be obtained. (Photographic reconnaissance is more fully described in the preceding section of this paper.)

From such information the likely paths can be laid out. As many checkpoints as possible are identified for the pilot's use in keeping on the line during the profiling run.

Control

Ground elevations of the terminal points of flight runs must be obtained to a good degree of accuracy, since other points on the path are referenced to them. This may be done by ground methods (altimetry or levelling from the nearest control elevation),or by radar profiling from the nearest control point. Stereo methods applied to reconnaissance photographs in certain cases may give the required elevation with reference to a known point with sufficient accuracy, provided the distance between the points is not too great.

The techniques of radar profiling may arbitrarily be divided into high-altitude and low-altitude methods. The latter is currently being used for a good deal of microwave survey work, with generally satisfactory results.

Advantages

Radar profiling offers the advantages of all aerial methods over ground methods, namely speed and suitability for terrain where access by foot or vehicle is difficult. In comparison with other aerial methods, it offers one of the lower cost ways of obtaining profile data for microwave paths, provided that the number of possible route choices can be sufficiently narrowed by other means. A rough cost estimate of profiling by this technique is of the order of $20 per final path mile, allowing for a maximum average of three separate routes per path to be profiled. Where more or fewer routes per path need to be flown, the cost will vary accordingly.

Weather conditions are of importance in the use of this method, apart from weather which would prevent flying altogether. Although the radar signal itself is relatively unaffected by atmospheric conditions, the barometric altimeter to which all readings must be finally referenced may be very much affected. During a period of unstable barometric conditions, it may be necessary to wait until the disturbances have passed before attempting to proceed with profiling.

This is a rapid method of obtaining data, which is available in usable form within a relatively short time after the flights are completed. This permits prompt evaluation of the results, so that additional paths may be profiled if required while the plane and crew are still in the area.

Availability

To do radar profiling work, an airplane must be specially equipped for the job, a rather expensive process. For example, a typical installation might include two precision radar equipments, two sensitive no-lag type barometric altimeters, synchronizing and recording equipment (photographic, tape or other) to make simultaneous records of all instruments, a counter, and a magnetic tape for recording flight commentary time-correlated with the other information. One or more aerial cameras are also generally carried.

At the present time there are not a great number of planes so equipped, but if the demand justifies it, others can be prepared. The problem of obtaining sufficient technically trained personnel willing to fly at low altitudes is not inconsiderable, but it is not expected that this will limit the availability of this method for microwave purposes.

Limitations

Accuracies in profiles obtained by this method are limited by the skill of the crew in making and interpreting measurements, difficulties in flying over certain types of terrain, dependence on accurate barometric information, and problems in flying a straight line due to mapping deficiencies and other factors.

Procedure

In low-altitude radar profiling, the actual profiling run is usually preceded by a higher altitude (above 1,000 feet) reconnaissance flight to spot possible sites and plot tentative flight lines; also to observe check points (such as road intersections) on the line. This is followed by the low-level profiling runs, at altitudes of 50 to 400'. During this flight the pilot flys as straight a course as possible from check-point to check-point as previously established. Where these are too widely separated, the crew may drop flour bags or other identifying marks along the flight path, to attempt to establish repeatability of course on return runs. Results on any path are generally not accepted by the crew until they have been duplicated by at least one corroborating profile on a subsequent run.

Equipment Data

The radar altimeters generally operate at a fairly high frequency and use directive reflectors in order to minimize beamwidth. One altimeter[1], for example, operates on a wavelength of 3.2 cm., and uses a four foot parabolic reflector to achieve a 1.5 degree beamwidth. This particular equipment has a crystal oscillator (for stability) as the heart of the timing circuit to control the firing of the transmitter and initiate the timing cycle. A pair of movable gate pulses straddles the video or echo pulse in a symmetrical fashion. If the gate pulses are forced to move to maintain symmetry, then their motion controls the charge on the grid of a tube in whose plate circuit the coil of the recording meter is located. The change in charge on the grid controls the plate current which is then recorded by the meter. Automatic sweep and automatic

tracking circuits enable the gate pulses to search for and lock symmetrically over the video.

Calibration of the altimeters is generally done in the following way. After the radar altimeters have been warmed up, they are calibrated in the air, using a standard delay line. Then a level flight pass is made over a flat area of known elevation, such as a lake, and the pressure altimeters are set to read correct elevation of the airplane above sea level or other reference, based on the readings of the radar altimeters and the known elevation. The instruments may then be checked by flights at different altitudes over the known elevation.

Some of the radar profiling systems (like the one described above) provide substantially continuous recording of ground elevations. Others obtain discrete readings at sufficiently frequent intervals to provide a satisfactory profile. An instance of this is the method in which the altimeters and other instruments are photographed, e.g. at one second intervals. Although this represents distances between readings of the order of 150 feet (at 100 miles per hour), any unusual elevation changes between readings would be noted in the flight commentary for subsequent checking. In addition, the return profiling run may provide a partial check on this, if the second set of reading are out of phase with the first set.

In most radar profiling procedures the information obtained must be correlated and adjusted after the flight is completed in order to obtain the final profiles.

Accuracy

Independent tests[2] conducted to determine the suitability and accuracy of one radar profiling method yielded the following information:

1- The accuracy of determining elevation data at the point measured is generally within ±20 ft., with maximum deviations per path generally not exceeding 60 ft. (When this method was first used, one source of errors was the lag in barometric altimeter readings when the airplane changed altitude rapidly. This has been fairly well overcome by use of a sensitive aneroid altimeter which provides a current or voltage to operate a recording meter without mechanical loading on the aneroid).

It should be noted that the pertinent elevations on the profile are usually the possible obstruction points, which are generally few. Therefore, even rather large errors in the recorded elevations of lower points may be of no important consequence in evaluating the path for microwave purposes.

2- The precise determination of the positions of terrain features is a more difficult matter for several reasons. The planimetric mapping available in many areas is not particularly reliable (so-called county maps, for example, often contain ¼ to ½ mile position errors) and therefore may not be very satisfactory for establishing the flight line. Another possible source of positional error is the difficulty of flying an airplane along a straight line, although with sufficient checkpoints and a competent pilot such errors may be kept within tolerable limits, provided that the terrain features do not exhibit rapidly changing elevations across the flight line. Variations in airplane speed may have the effect of apparently displacing terrain features along the line of flight, but from a practical point of view this is generally not important, since the amount of displacement is not great by comparison to the half-path length, the criterion in determining Fresnel zone clearances.

The low-altitude radar profiling method can usually provide the profile information needed for microwave paths with sufficient accuracy, provided certain precautions are taken. These include repeating the profile run in the return direction to obtain confirmatory data, allowing for possible changes in barometric conditions during measurement runs, and checking doubtful elevations (such as ridge lines rising rapidly above average terrain) by ground survey or other means. Foliage must be taken into account, since the radar may penetrate this either completely or partially.

High-Altitude Profiling

Another way of utilizing the radar altimeter for profiling is to make the readings from an altitude of the order of 5000'. This procedure has been used mainly for providing sufficient vertical control data for use in conjunction with the photogrammetric method, and has not been used to any extent for microwave applications.

Although the elevation errors average less than 20 feet, random errors in rugged terrain may exceed 100 feet. An even more serious difficulty from the microwave standpoint is that the plane cannot fly a straight line within acceptable limits,

and therefore the profile obtained does not represent the microwave path desired.

These difficulties do not detract from the use of this method in its current major application, providing vertical control for use with photogrammetry, since only selected points on the profile, generally of smooth surface, are used, and these are correlated with the plane's position at the time by photography. Whether these points are on a straight line or not is of no importance to the photogrammetrist, as long as they can be located with respect to horizontal control data.

One problem experienced by those using this method may be of interest, in the light it throws on the use of the pressure altimeter. Although the radar altimeter readings are corrected with respect to the barometric altimeter as the plane varies in altitude, the barometric altimeter itself is subject to a peculiarity of the atmosphere. "If there is no movement of the air mass, the plane can fly a level course by maintaining a constant reading on a sensitive altimeter. If there is movement of the air mass, to fly such a level course, or to reduce the results of maintaining a constant pressure altitude to such datum, requires a pressure correction somewhat similar to correction for tide. As the profile lines are flown, a record of the amount of 'crab' necessary to keep the plane on its designated course is kept, and from this a correction is worked out for the resulting rise or fall of the plane due to keeping to an 'equal pressure' altitude"[3].

Path Testing Using Helicopters

The basis of this method is the ability of helicopters to hover almost motionless. To determine required tower heights for a proposed path, two helicopters equipped with microwave transmitting and receiving equipment hover over the proposed sites. A curve of signal strength versus height from ground can be plotted from the no-signal point to and beyond the first Fresnel zone. Preliminary reconnaissance, possibly by the helicopters themselves, is required to determine which points are plausible site locations. (This is similar to the reconnaissance required to determine proposed paths as discussed in the radar profiling method.)

In one version of this method[4], one helicopter carries a microwave transmitter radiating horizontally towards the second in a 30° beam. The other carries receiving equipment whose antenna beamwidth is 8°. The received signal is observed visually on a cathode ray tube and aurally in headphones. The visual indication permits determination of relative field strengths and estimation of Fresnel patterns. The aural signal enables the pilot to maintain proper heading. Communication between participating helicopters is provided by a separate radio-telephone system.

To enable the helicopters to hover into the wind at all times, it is necessary to provide 360° rotation of the antennas. This is accomplished by having two antennas available on each helicopter, each providing slightly more than 180° of rotation. The choice of antenna depends on the direction of the wind at the time of operation.

The height of the helicopter at any time may be read on a survey-type portable barometric altimeter. The helicopter is lowered to the ground, where the altimeter is set at zero. Thereafter, heights can be read directly to accuracies within 5 feet. Plumb lines may also be used for this purpose, as well as for determining tree heights.

Experiments have shown that the signal cutoff point is quite sharp as the beam falls below the grazing point. To take readings over any site takes only a few minutes.

The frequency used when making the tests need not be the frequency of the proposed link. In the particular case under discussion, a frequency of the order of 9,000 mc was used, because the equipment was smaller and lighter for the same antenna patterns, and was readily available. Adjustments for differences in Fresnel zone radii at different frequencies can readily be made.

A variation of this method is to use light sources in the helicopters. Heliostat type equipment when there is sufficient sunlight, or very strong focussed lights at twilight or at night, will provide information on the elevations required to achieve line of sight. Proper Fresnel zone clearance requirements can then be established based on this data.

A further variation of the method would be the use of one helicopter with a portable tower replacing the second helicopter, or even a balloon carrying a light over the tentative site location.

Cost of a helicopter plus pilot varies considerably with the amount of utilization which can be scheduled, and

may be expected to be between $50 and $100 per hour.

The helicopter method may be very advantageous in terrain which would be difficult to profile, for reasons such as heavy vegetation cover, few control or check points, and the like. In addition, in areas where interference possibilities exist with other systems, path testing provides a direct method of determining the amount and nature of the interference.

It is important to recognize one inherent pitfall in this method. Since microwave signals of all frequencies are subject to refraction, and since the index of refraction varies with atmospheric conditions, signal strength patterns recorded at one time will not necessarily represent patterns to be expected at another.

Application of Methods

The application of the various methods of aerial survey described above to the problem of microwave path location depends upon a number of factors.

Mapping

The first of these, and probably the most critical factor, is the status of the existing mapping throughout the area of concern. If the area is completely covered with large scale topographic maps of good quality, a large part of the physical survey work is already done. The microwave engineer has all ready for him a record of positions and elevations on which to base his studies. He has only to survey the remaining items of tree heights and man-made obstructions at critical points along the paths he studies. At the other extreme there may be no topographic maps and the planimetric maps may be nothing more than pictorial in nature. Positions and elevations may need to be surveyed in this case for many points throughout the area.

Nature of Area

The second factor governing application of methods is the nature of the area involved. This may vary from flat and heavily forested to barren and mountainous, and with many different degrees of cultural development in the areas. These influence greatly the need for varying degrees of survey precision and accessibility to the areas for survey.

Communications Requirements

The third factor involves the nature of the communications system being studied and the facility for which it is being planned. Although the prime subject under consideration here is survey for microwave installations, this can seldom be considered by itself. The interrelationship of microwave to other requirements such as VHF coverage required or connection to pole lines may require survey methods suited to several needs. In addition, the location, number and separation of designated points of communication may affect the choice of survey method.

In addition, it may occur that the communications system is being planned for some facility such as a pipe line or power transmission line which is itself still in the planning and location phase. In such a case the surveys required for location studies of the facility may also serve partially or fully the requirements of the microwave survey.

Time Schedule

The fourth factor is the time allocated to the survey and system of the system involved. Related to this is the season of the year and the section of the country involved, since these will affect the time required for survey by the different methods. The availability of each procedure will also enter here. The time element is usually set by the ultimate user of the communications system, and he often sets a time too short to permit complete surveys to be made.

Cost Considerations

The fifth factor is that of cost. Normally it should be expected that the most economical survey method that provides the required results would be employed. However, the time element in many cases may take precedence over the cost factor, since the total survey cost generally represents only a minute fraction of the overall cost of the facility and of the revenue which will be realized when the facility is put into operation. Therefore, more costly procedures may have to be used to meet the schedules imposed.

Special Factors

In special situations, such factors as the following may be of importance:

Effect of survey method on subsequent site procurement (when people have advance notice of something afoot, prices tend to go up.)

Flexibility, to permit changes after completion of survey, for such reasons as

selection of a new site (if one originally chosen were unobtainable), revised system requirements, and the like. In such cases, availability of needed additional information with minimum time and cost would be desirable.

Choice Criteria

It should be evident from the nature of the factors affecting selection that no easy criteria may be set forth to guide the choice of survey method. Proper selection may best be made by an engineer with training and experience in the fields of communications, topographic and geodetic surveying, photogrammetric mapping and path testing techniques. It is unlikely to find such a combination of qualifications in one individual. It is more likely that a communications engineer and a topographic engineer will jointly provide the desired combination. Even though it is felt that only such engineers should make the choice after knowledge of the governing factors is established, some indication of how such a choice is made may be obtained by pointing out some of the advantages and disadvantages of the various methods under differing conditions.

Although it is not the purpose of this paper to discuss methods of ground survey, this subject cannot be entirely disregarded in discussing aerial methods, since there is no case which does not involve some survey work on the ground. Similarly, ground reconnaissance cannot be entirely disregarded.

Let us now examine the methods previously treated in terms of their application to microwave work, considering the purpose, advantages and weaknesses of each.

Reconnaissance - General

As in any location problem, reconnaissance is the first phase. It involves a number of steps, usually beginning by assembly of all existing data for the area of concern. First reconnaissance is generally qualitative in nature, followed by further reconnaissance to obtain more or less rough quantitative data, after which another method is used to prepare final designs.

Reconnaissance - Visual

As implied by its name, visual aerial reconnaissance finds its application during the reconnaissance phase. It has several significant advantages. It is low in cost and is readily available, since small plane transportation may be rented at most airfields at very reasonable rates. No special equipment need be installed in the plane. It is probably the fastest way to get a general impression of unmapped or poorly mapped terrain, and is usable under most weather conditions.

Two obvious disadvantages of a visual reconnaissance are that it produces only very approximate quantitative data, and that it produces very little in the way of a record of its work. Such record is limited to written records on map or paper and the few photographs that may be taken on the flight. Its application is believed to be limited to the following cases:

a- In unmapped areas where existing data gives no general information of the nature of the topography and culture.

b- In areas well-mapped topographically, where more general information may be needed on vegetation and culture not usually shown on the map.

c- In well-mapped areas where tentative paths have been selected from the maps and where a check is needed along critical points of the paths for possible obstructions, reflecting bodies, facilities at designated sites, and like information.

Reconnaissance - Photographic

The aerial photographic reconnaissance method has two principal advantages. The first is the complete record of the great amount of information it brings into the office for subsequent study; the second is the relatively good approximation it provides in determining horizontal positions for features of concern.

Offsetting these are three principal disadvantages. First, the cost of securing photographs is more than nominal, and a plane must be specially equipped with camera and experienced operator to make these photographs. The second is that the effects of weather may produce some delay in getting underway with the work. This may be partially offset by the fact that once the photographs are obtained, the office phase of the reconnaissance may proceed independently of weather conditions. The third drawback is the relatively low accuracy in approximating elevations for features of concerns.

The application of this method will generally be confined to those areas where there is no topographic mapping, or where existing topographic mapping is very poor. When preliminary studies or reconnaissance

indicate a need for a fairly wide exploration of such areas, this method will usually provide the cheapest and fastest means.

From the photography obtained it may be possible to select relay station sites and obtain sufficient path information to lay out a system on a tentative basis. It is easier to pin down the likely paths in rugged or mountainous terrain than in rolling or relatively flat terrain. High points can be easily picked out in rough terrain and their locations identified fairly accurately, since photo-reconnaissance gives good position information. In rolling terrain or any terrain where changes in elevation are gradual, it is more difficult to identify high points and therefore to make reliable path selections. While some field checking would be required in either case, there will generally be fewer critical points to check in rough country.

Survey Methods

The other methods described are applied in the final path determination phase of the work, after the reconnaissance has been completed. Radar profiling and photogrammetry would be applicable principally in areas for which no suitable topographic mapping exists, since they produce essentially topographic data. In such areas the choice of either of these aerial methods must be weighed against ground survey methods.

Radar Profiling

The radar profile method is particularly advantageous to use where reconnaissance has narrowed the choice of possible paths requiring measurement, and where accuracy of horizontal positioning of profiles is not too critical. Although no fixed dividing line can be set, radar profiling is likely to be economical up to the point of measuring three to four alternate routes per path. Beyond that it is likely to become too costly, since the cost of this method is largely a function of the length of line measured.

Within these limits the method is fast, and it produces data quite quickly upon completion of the flying phase of the work. However, the flight runs may be delayed by adverse weather because of the dependence of the method on stable barometric conditions.

It should be noted that while the radar method is subject to weather delays, the weather factor is less critical than is the case in the photographic phase of the photogrammetric method, for example. In using the radar method, satisfactory weather conditions may be expected on the order of 50% of the time, while weather conditions suitable for making the photography for photogrammetry may occur on the order of 10% of the time.

Tests have indicated that the method gives satisfactory accuracy in elevations but that much care must be exercised to secure such results. The method is weak in determining horizontal positions, however, since it is dependent upon map spotting and/or navigational procedures. In many cases this weakness is of no concern, but it must be taken into account for those cases where it is critical.

Before radar profiling method can be applied, the tentative sites must be designated in such a way that positive identification by the flight crew can be made, and elevation data for these points must be furnished as a basis for the profile. This data must be obtained from some other survey method, such as ground altimetry techniques from the nearest control point.

Photogrammetric Survey

The photogrammetric survey method finds its application where a number of lines or points spread over a considerable area must be surveyed. The costs of this method are largely a function of the number of photographs handled. Since the photographs cover a considerable area, survey data can be produced throughout that area for about the same cost as along any single line through the area. The method produces the data rapidly once the initial lags are overcome, and is completely independent of weather at that point since all work takes place in the office once the photography is obtained. This can be an around-the-clock procedure if necessary.

The accuracy of both elevations and horizontal positions is satisfactory for final path determination. Sufficient information concerning sites can be obtained, up to the point of determining subsurface conditions and cost of purchase of the land. The aerial photography used in the method is also suited to the earlier reconnaissance phases. Photogrammetry may also provide survey data for other related purposes, such as VHF coverage, construction and access maps for tower locations, or maps for right-of-way procurement. This additional data may be provided at very small cost additional to that required for the path determination.

The principal disadvantage of the

method is its relatively high cost. As pointed out earlier, the cost becomes favorable only when a considerable amount of survey data must be obtained. Under some combinations of area and season, the effects of weather on securing suitable photography may be a disadvantage.

The path testing technique using helicopter methods finds particular application in terrain which is difficult to profile, due to factors such as heavily wooded areas or deficienty of identifiable check points. It may be used in any terrain after a preliminary reconnaissance has indicated a group of plausible sites.

It has the considerable advantage of speed, being faster than any of the profiling type methods because the engineer obtains information concerning obstruction characteristics of potential paths without actually flying over them. In addition, optimum tower heights may be determined within close limits, based on the signal strength variation with altitudes of the aircraft.

The method has several weaknesses, chief among which are the following: the microwave signal is susceptible to refraction, and since refractive conditions change from day to day, clearance conditions obtained at one time cannot be taken as wholly representative of clearance at other times; hovering sufficiently still with the helicopter poses problems; this method of itself does not provide data concerning the path which may affect propagation, such as water bodies, tree growths, etc.

REFERENCES

1. Photographic Survey Corp. Ltd., Toronto, Ontario, Canada.
2. Conducted for RCA Victor Division, RCA, by Frank H. McIntosh, Washington, D.C.
3. Report on Airborne Profile Recorder, by Harry T. Kelsh, U.S.G.S. Washington, D.C., printed in Photogrammetric Engineering, March, 1952.
4. Based on material in the article "Siting Microwave Antennas", by B.I. McCaffrey, PSC Applied Research, Toronto, Canada.



Fig. 1
Aerial photographer with camera and view finder installed in an airplane.

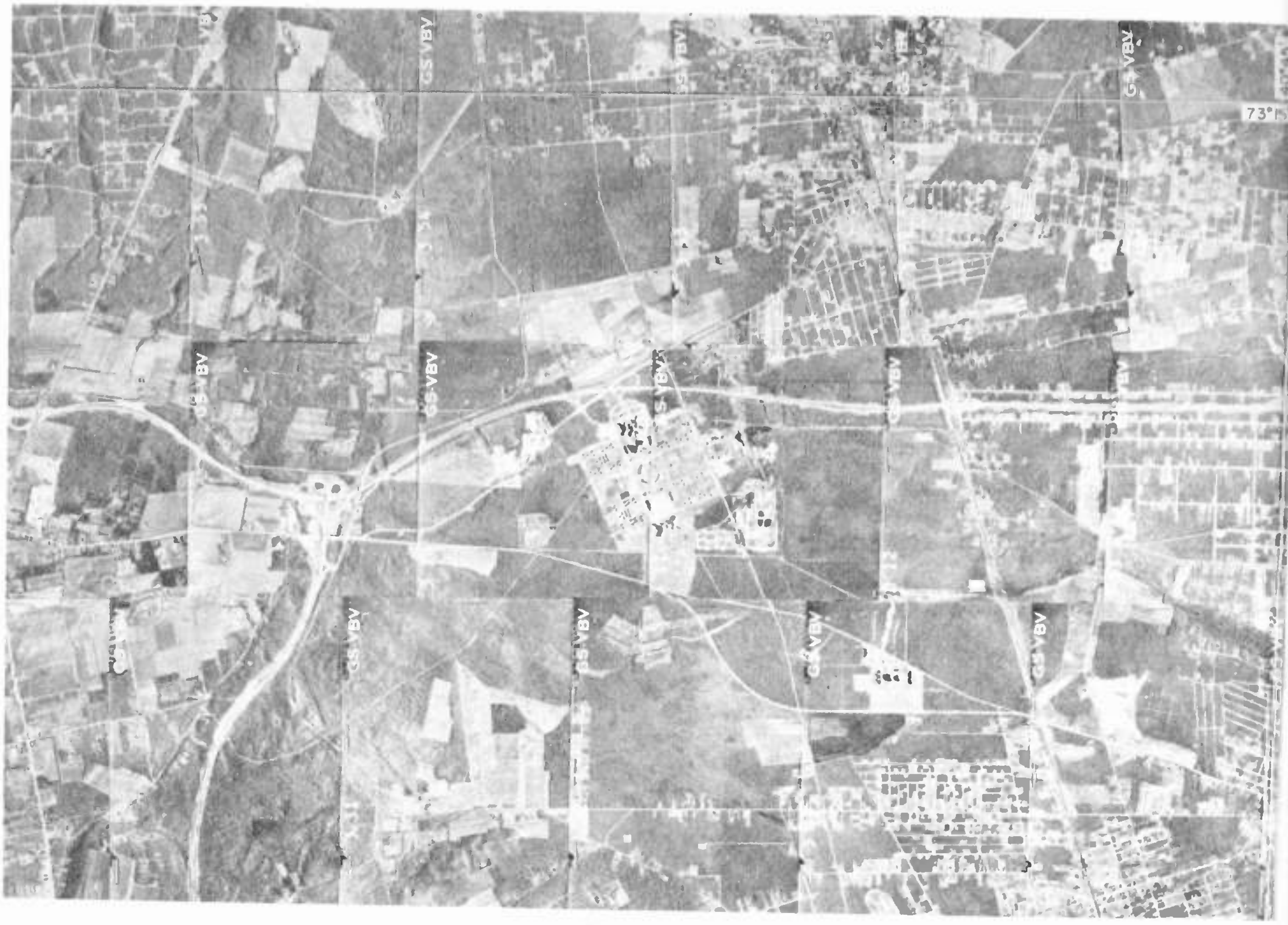


Fig. 2
A portion of a photographic index.
This reproduction is approximately 30% of the scale of the original aerial photographs.



Fig. 3
A small portion of an enlarged aerial photograph covering an area in west central Wisconsin.
Scale of original photograph 1:60,000. Scale of this reproduction 1:15,000.

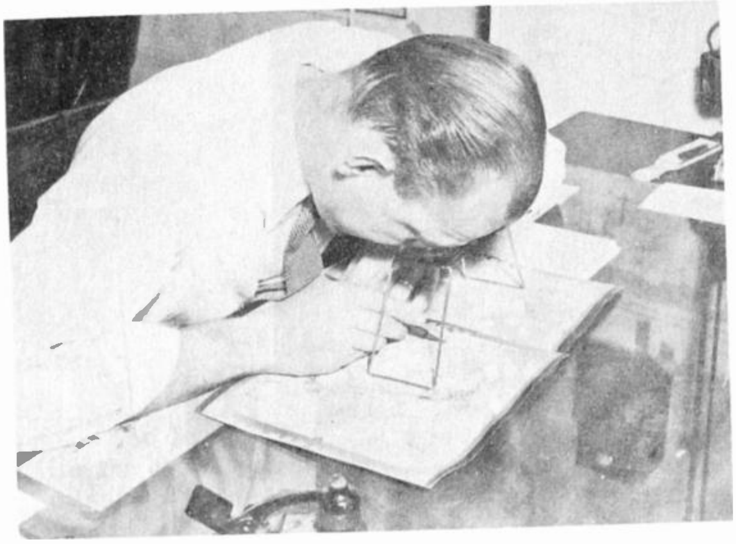


Fig. 4
Folding pocket stereoscope being used for study of aerial photographs.

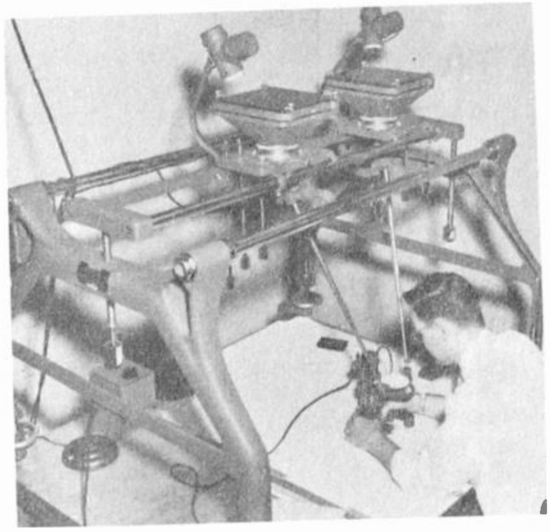


Fig. 6
A Kelsh Plotter. One of the simpler photogrammetric instruments used for precision aerial surveys.

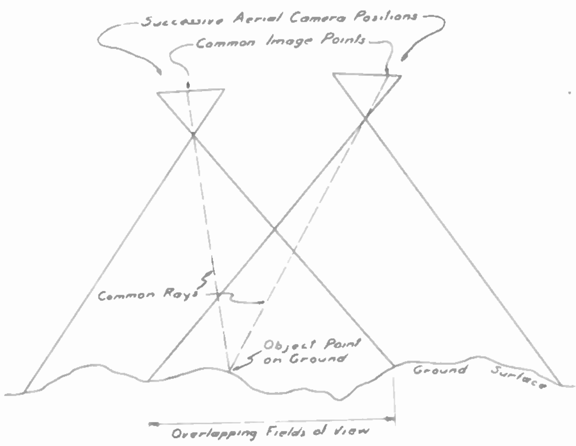


Fig. 5
Sketch illustrating principle of photogrammetric surveying. View is in elevation on vertical plane containing the two aerial camera positions.



Fig. 7
A Wild Autograph, Model A-5. One of the most complex and precise photogrammetric instruments.

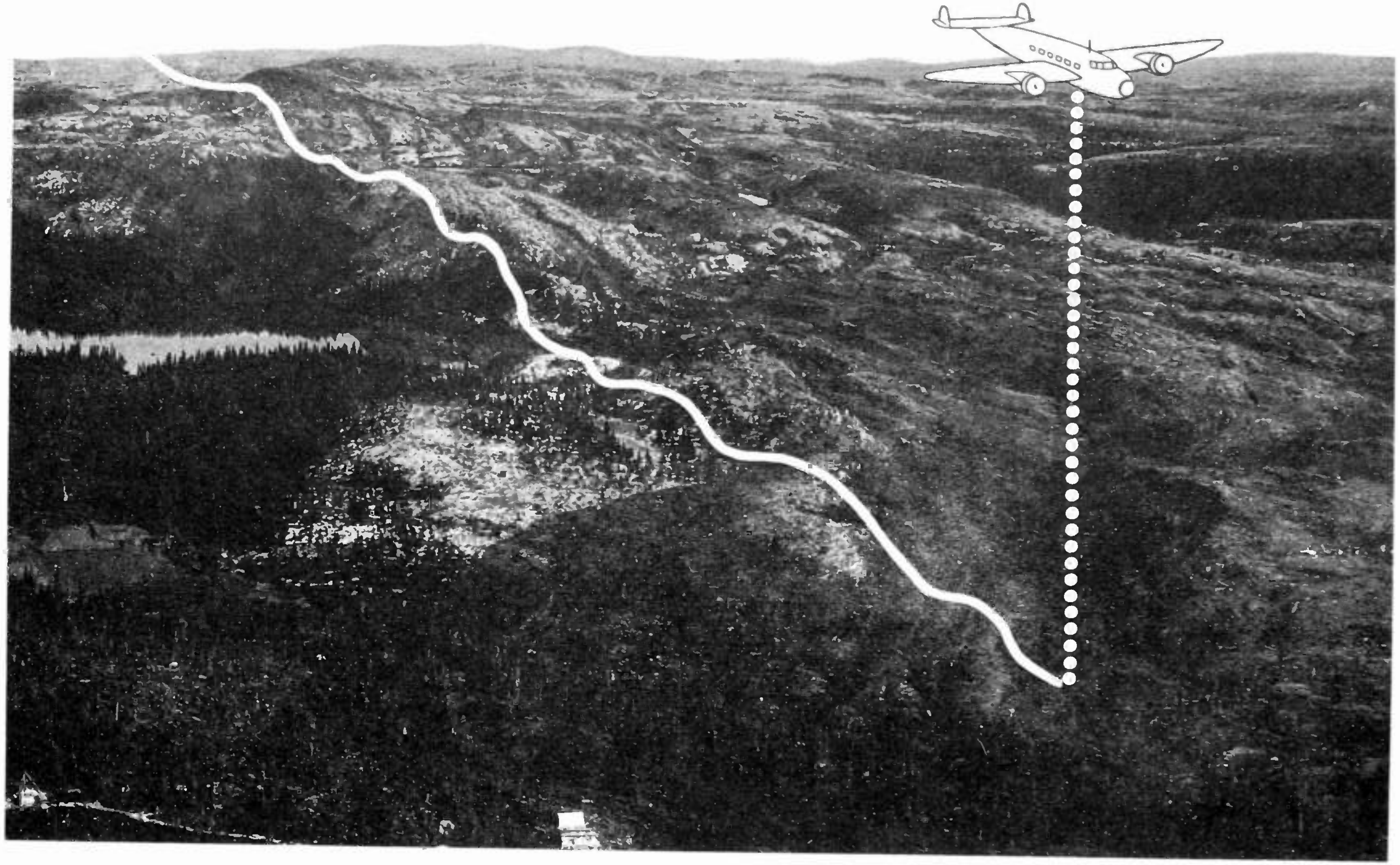


Fig. 8
Radar profiling in schematic representation. Photo by P.S.C., Toronto.

THE DEVELOPMENT OF A PRODUCTION RADOME TESTER

Robert P. Walcutt

Antenna Research Laboratory
Thompson Products, Inc.
Columbus, Ohio

SUMMARY

The development of a test instrument for the checking of reflection and transmission coefficients of radomes in production is described. Various investigationns were made to determine by what means interaction between parts of the RF system could be minimized or eliminated.

The results of these studies are incorporated in the equipment which has been constructed. The equipment is automatic in operation, once initial adjustments are made. It is well suited for use in manufacturing plants as a quality control device.

The purpose of the investigation is to devise and build a system for electrical evaluation of radomes. The qualities to be evaluated were defined as "transmission coefficient", the amount of energy reaching the target area through the radome divided by the amount of energy reaching the target area in the absence of the radome, and "reflection coefficient", the amount of energy returned to the radiating system due to reflection from the radome divided by the total outgoing energy. No requirement for measurement of boresight error was included, and the system devised includes no means for this measurement.

Five specific requirements were added to the general problem. The system and equipment incorporating it should be adaptable to measurement of production radomes (quality control); the equipment could be operated by unskilled, but not untrained, operators; a visual indication of the information should be provided while obtaining a permanent record; the system, as constructed, should be accurate to within ± 1%; the equipment should operate throughout the 2400-3700 mc and the 8500-9800 mc bands.

Measurements of transmission coefficient have been carried out under some difficulties in the past. It has been necessary to make two distinct measurements, with the transmitting antenna moved through a quarter wavelength between measurements, and to average the values obtained to get an evaluation of the radome. This procedure has been necessary due to the large interaction between the test specimens and the micro wave equipment in the test setup. Interaction between the test antennas has required that measurements be made over relatively large ranges.

Cognizance of the general requirements, specific requirements, and the past history of the practice indicated that certain basic studies were necessary before any construction could start. To facilitate these studies, the services of Sinclair Radio Laboratories, Ltd., Toronto, Ontario were obtained. Sinclair Radio Laboratories carried on the major portion of theoretical work which resulted in a current practical equipment for measurement.

The measurement of transmission coefficient by the simple expedient of two measurements of field strengths, one with the radome in place and a second without the radome, has been impossible in the past because the outgoing power level varies greatly as the disance from the radome to the antenna changes. This variation is cyclic with distance and the period depends on the wavelength of the test frequency. It is caused by interaction among RF components. Three major sources of interaction are experienced in the radio fre-

quency equipment. These exist between the radome and (1) the generator, (2) the transmitting antenna, and (3) the receiving antenna.

One of the first approaches considered to elimination of this power variation was that of providing a continuous mechanical shift of the transmitting antenna with respect to the radome. The shift would have to be through several half-wave lengths in order to obtain a true average value and occur quite rapidly so that the measurement equipment would indicate only the average rather than the individual variations due to movement. This system would actually be a variation of the old test system whereby two individual checks were made. It was discarded as a possibility after consideration of its mechanical requirements.

Another system considered was that of shifting the frequency continuously across a small portion of the spectrum centered about the nominal test frequency. This would serve to cause the antenna to radome spacing in wavelengths to vary and accomplish the same thing as the mechanical shift. Several disadvantages to this are apparent, the more important of which is the difficulty in obtaining generators capable of this swing with reasonably constant output.

A system considered which provides the most satisfactory method of measurement is one in which the radio frequency system is specifically matched to perform the operation needed.

Under usual bench conditions, a more or less arbitrary and inexact match between the generator (klystron) and the micro-wave plumbing results. This match may be the best observed "maximum power transfer" condition. Any power travelling back towards the generator as a result of some discontinuity will, in general, encounter a second discontinuity when it reaches the junction of the generator and plumbing. At this second discontinuity a power split occurs wherein some of the power is transferred into the generator and dissipated in the internal impedance and the remainder travels along with the outgoing power. Obviously, the phase of this reminder is important when added to the generator output to yield total outgoing power. Thus the outgoing power level is a function of (1) generator output (2) phase and magnitude of any reflected power, and (3) the character of the mismatch at the generator-plumbing junction. Since a radome represents a discontinuity in the path of the outgoing power causing the reflected power to be of variable phase and magnitude as different parts of the radome are illuminated, the resultant power illuminating the radome will vary.

However, under a condition where the micro-wave plumbing sees an exact match looking back into the generator all power reflected from any discontinuity will be absorbed by the internal impedance of the generator. Since there is no re-reflection at this generator - plumbing junction, the outgoing power in the plumbing will remain constant.

A system for obtaining this exact generator to plumbing match has been devised. A short circuit is placed in the microwave plumbing and its position varied while the outgoing power level is observed by means of a directional coupler placed between the matching system and the moveable short circuit. This outgoing power will vary as the short is moved except when an exact match is obtained. When this proper match is realized no change will be noted in outgoing power.

This power must be radiated from an antenna within a radome. Energy reflected from the radome will strike this antenna. At best, an antenna can absorb only half the energy impinging upon it, the rest being scattered. Some of the scattered energy will combine with initial outgoing radiation and thus cause a change in the energy level which illuminates the radome.

The only matching arrangement which will prevent the scattered energy from causing trouble is one in which the antenna impedance is only reactance. Because the absence of a resistive component clearly indicates a lack of radiating ability the

antenna is useless as a transmitter. Thus we see that no matching condition at the antenna can eliminate the interaction with the test specimen. This interaction effect is increased by any two of three conditions; a test specimen of high reflection coefficient; a large area of the test specimen normal to the direction of propogation; a high gain transmitting antenna. Fortunately, radomes do not in general fulfill the first two requirements.

Since radomes are usually of compound curvature throughout and specifically designed to have low reflection coefficients the fact that a reasonably high gain transmitting antenna may be used will not in itself cause any serious trouble. If it is desired to test flat panels some sort of isolation between the transmitting antenna and the test specimen must be provided. A pad of lossy material which will provide approximately 10 db. loss in the transmission path is sufficient to isolate the two. This pad must be such that it is essentially reflectionless and so oriented that any reflections which do exist are not allowed to return to the transmitting antenna or the test specimen. In practice this pad has been composed of Celotex sheets with lossy coatings of Aqua Dag applied. Other forms of the pad have been tried and found successful.

Interaction with the receiving antenna can be eliminated by using a very long range or by use of a very small aperature receiving antenna. The latter seems to be preferable. Reradiating properties of the receiving system are reduced to a minimum by placing a very small lossy sheet in front of a very small receiving antenna.

If the test setup recognized and properly compensates for all the difficulties as described, a useable arrangement can be had for making single readings of transmission coefficients.

The case for a single reading reflection coefficient measurement is more complex. Even under the ideal match condition described, the amount of power measured as returning toward the generator will vary as the antenna-radome spacing varies. This is not thoroughly understood but may be considered in light of the fact that the radome is definitely in the near field of the antenna and its movement corresponds to a change in the complete antenna system. Some relatively impractical systems for isolating the radome from the antenna by loss have been conceived but can not be put in practice.

The amount of reflected power in the RF plumbing is also a function of the transmitting antenna itself. This creates a new set of circumstances for reflection coefficient measurement. If results of any validity are to be obtained it now appears that the radome should be tested in a set up simulating the actual application and that the transmitting antenna should be identical with that to be housed by the particular type of radome under test.

Under these simulated conditions reflection coefficient measurements will yield a practical answer from an application standpoint in that it is now possible to learn (1) the general level of RF power returned to a radar system by the radome, and (2) the rate of change of RF power returned. These are important in application, since in the first case the operation of the magnetron will be affected, and in the second the performance of any AFC system must be considered as it compensates for magnetron pulling.

A mechanical-electrical system was devised to support a transmitting antenna, the typical antenna for the radome under test, with an arrangement for rotating a radome around this antenna. A 0° to 360° rotation is accomplished during which time a plot of transmission coefficient is made. Automatic circuits switch the equipment to measure reflection-coefficient and this measurement is accomplished during a rotation of the radome from 360° back to 0°. This provides a complete set of data for that sector of the radome scanned. The equipment then automatically tilts the radome and its axis of rotation 15° while holding the transmitting antenna in a fixed position and again starts plotting transmission coefficient.

Thus the radome is sampled in sectors 15° apart; this is felt to provide sufficient information for production quality control of radomes. The scanning may be stopped at any angle of tilt by presetting a selector switch.

The equipment is provided with manually operated controls so that any function may be carried out independent of the automatic circuits.

Due to the length of time required to obtain a complete set of information on a radome it was felt necessary to provide servo driven attenuators in the RF system. These attenuators will hold the output power constant for long periods of time.

The receiving antennas for 3 cm and 10 cm are of smallest possible aperature to reduce re-radiation: an open guide in the first case and a fat dipole in the second. These antennas are placed flush with an absorbing screen which reduces reflections from objects beyond the test area. A lossy cloth is placed in front of the receiving antennas to further reduce the amount of energy reradiated to the transmitter. The receiving antennas are located 15 feet from the transmitting antenna.

Arrangements are to be included for the measurement of panels. For the purposes of comparison and calibration, standard test panels will be provided. In order that a realistic value for reflection coefficient can be obtained on these flat panels it would be necessary to measure them without the lossy pad. This requires the use of the former system whereby two measurements are made and the results averaged.

The RF system exciting the transmitting antenna consists of a Varian X-13 klystron at 3 cm and a cavity assembly from a Hewlett-Packard 616A Signal Generator at 10 cm. Each oscillator feeds into a 3 db pad, thence to a tuner which accomplishes the generator to plumbing match. Immediately following is the servo controlled attenuator followed by a directional coupler sampling outgoing power and feeding information to the servo. A second directional coupler provides signal for reflection coefficient measurement. Immediately preceding the antenna is another tuner which matches the system to the antenna under a no test-specimen condition.

The electronic equipment necessary for this system is reasonably straight forward. Klystron power supplies are electrically regulated to less than 3 millivolts of ripple. Klystron heater supplies are DC to prevent FM due to hum. The square wave modulator for the klystron has a very flat top so that no FM is introduced.

All amplifiers for measurement, monitoring, or automatic stabilization are selective so that only the fundamental component of the square wave is used. The signal amplifier is of a 15 cycle band pass type with the operating frequency near 1500 cycles. The last stage is a power amplifier and diode rectifier arrangement which feeds an 0-1 ma precision meter and a recording voltmeter.

The recording instrument is a modified Leeds & Northrup Speed-O-Max having azimuth angle on the X axis and signal amplitude on the Y axis. Included is special equipment for automatically marking each recorded chart with identifying information.

The design goal for accuracy has been 1%. Each component has been constructed with a zero tolerance in mind so that the overall cumulative error will be less than 1%.

This equipment and system of measurement will permit the rapid and accurate single reading measurement of transmission coefficient of radomes or dielectric panels. Reflection coefficient measurements of radomes can be made which produce an indication of the general level of the power which may be reflected back into a radar set and some idea of the rate of change of this power. Reflection coefficient measurement of dielectric panels can be accomplished accurately with this equipment by the two-reading system.

A CORRELATION DIRECTION FINDER FOR GUIDED MISSILE RANGE INSTRUMENTATION

Marvin S. Friedland and Nathan Marchand
Air Force Missile Test Center
Patrick Air Force Base, Florida

1. General Requirements

A study of the requirements for an electronic trajectory measuring system on a guided missile testing range discloses many conflicting and incompatible criteria. For a medium and long range, more than 200 miles, some of the criteria generally stated as desirable may be listed as follows:

a. A trajectory measuring system whose accuracy capability competes with that of a modern optical tracking system such as a cine-theodolite chain.

b. A single station system not hampered by the logistic support requirements for long base line systems, base lines of many miles.

c. A system not limited to line-of-sight characteristics of the ultra-high and microwave frequencies.

d. A system requiring a minimum or no electronic or other equipment in the test vehicle.

e. A system which is unambiguous, thus not requiring other trajectory measuring devices to assist it in making a determination.

f. A system not subject to the contamination of atmospheric noise or skywave reflections.

g. A system which does not require acquisition assistance to lock on the test vehicle once in flight, or does not require a large frequency spectrum.

An examination of these criteria leads one to conclude that the state of the electronic art has not progressed to the point where they may be met without compromise. Therefore, it is concluded that the systems in general use by the guided missile ranges have much to be desired in the way of improvements.

2. Equipment Limitations on Test Missiles

During the free flight testing of Research and Development test vehicles on a range, certain electronic equipments not generally related to the functioning of the vehicle are usually required on board. These are a radio telemetry transmitting system to provide data as to the internal performance of the vehicle and the associated end instruments to pick up this data at points of interest.

Therefore, to consider this equipment, as it may be also used for a trajectory determination function, would not entail the installation of additional equipment in the vehicle. The telemetry equipments in common use have been standardized and would not, in their present form, impose a serious limitation on their use for this additional function.

3. Direction Finder on Telemetering Frequencies

The use of a precision direction finder on the telemetry signal from a vehicle in test flight appears practical if special precautions are observed. The FM-FM telemetry band from 215 to 235 Mc is in a part of the spectrum where the atmospheric noise and sky-wave effects are small; further, the propagation characteristics at these frequencies have been explored and data is available. While it is true that such a direction finder would be limited to the usual UHF ranges, it appears that it could be developed into a practical equipment to be used at small down range stations.

4. General Antenna Theory

Basically, all direction finders work upon the principle of measuring the difference in time of arrival of an incoming radio wave (between different parts of the same antenna, when a single antenna is used, or between different antennas when an array is used). For instance, two antennas, I and II shown in Fig. 1, are separated by a distance D. A wave front arriving at an angle to the line of the two antennas will cause the voltage in I to lag behind the voltage in II by the number of degrees represented by the distance Δ where

$$\Delta = D \cos \theta \quad (1)$$

This can be expressed in electrical degrees, ϕ, where λ is the wave length.

$$\frac{2\pi\Delta}{\lambda} = \frac{2\pi D}{\lambda} \cos \theta = \phi \quad (2)$$

The first limitation in accuracy, using this method, is that the distance D must be very

large for the accuracies necessary, but with a very large D there is a multiple of solutions inasmuch as the system cannot distinguish the difference between ϕ and (ϕ + n 360°) when measuring. This must be considered because of the accuracy limitations of present day phase measuring devices, and signal to noise ratio. To resolve this ambiguity, it is proposed to use a very coarse system with the antennas spaced a distance $\lambda/2$ apart, and then use two sets of antennas spaced d and D apart as shown in Fig.2. Here the antenna array with a spacing d is chosen so that the very coarse antennas (which have no ambiguities) resolves the ambiguities in the other antennas. Similarly, the antenna array using a separation d resolves the ambiguities of the array using a separation D.

Crossed groups of arrays as described above are used for 360° coverage and for resolving elevation. A plan layout of the array is shown in Fig. 3. Actually, D is many times larger than d, the size being determined by the accuracy of the phase measuring equipment and signal to noise ratio.

The actual number of arrays to be used and their spacing must be determined by the desired system accuracy.

5. Phase Measuring Receiver

The problem is to measure the difference in phase between two complex FM signals.

The saving grace is, of course, that the two signals are exactly alike except for a delay between them. The receiver must be able to discard all extraneous information and to use all of the energy in the wave just to determine this phase difference effectively.

A cross correlation receiver is shown in Fig. 4. This receiver consists of four channels, A, B, C, and D, where phase comparisons are made between A and B, and between C and D. Except for a common local oscillator, each channel is a standard receiver up to and including the intermediate frequency amplifiers. A common local oscillator operating on a frequency 30 Mc lower than the incoming signal is used so as to eliminate any phase difference due to this oscillator since it will be cancelled out by being introduced equally into all channels. Isolation amplifiers are used in the oscillator circuit to prevent interaction between channels. It is important that the oscillator and RF tuned circuits track well enough so that no internal phase shifts between channels will take place. The RF and IF bandwidths are kept wide enough to make this point less critical.

Cross correlation is employed after the IF strips. A low frequency f_1 crystal oscillator is used for the cross correlation oscillator. For the signal in channel B to be compared with the signal in channel A, the f_1 signal is mixed with the local oscillator. This results in a f_1 higher signal being fed to the B channel mixer. The result is that the local oscillator signal being fed directly to the A channel mixer producing a 30 Mc difference frequency containing all the modulation and phase components and that the local oscillator plus f_1 is fed into the B channel mixer, producing a 30 + f_1 Mc difference frequency with similar components in this channel.

Mixing the signal from the B channel IF with the signal from the Channel A, IF will yield an f_1 signal containing only the difference signal between channel A and Channel B. Inasmuch as the only difference between the signals in channels A and B is the phase difference due to antenna displacement, the output will be an F_1 signal displaced in phase by that phase difference from the reference oscillator.

The f_1 IF is very narrow since the flow of phase information is very small and changing comparatively slowly. Thus, the phase information will be passed while the modulation on the signal will be rejected. A phase comparitor can now be used to compare the phase of the f_1 signal coming out of the IF with the f_1 crystal oscillator phase.

Any change in phase of the f_1 oscillator will automatically cancel out. The reason for using a crystal oscillator is to allow the use of a very narrow band f_1 IF. The phase comparison between the signals of channels C and D is made the same way as that for channels A and B.

6. Receiver Theory

It is possible to cross correlate two signals by mixing them together in a linear mixer (linear so far as output is concerned) and then passing the output through an integrating or low pass filter. By varying the time displacement between the two signals, the amplitude of the output will be the cross correlation function.

However, it is usually much more expedient to use the band pass equivalent of this method. To do this, one of the signals is displaced in frequency an amount equal to the final IF frequency. When the two signals are now mixed, the IF output signal has an amplitude which is equal to the cross correlation function, and which contains all of the phase information caused by cyclic displacement. Integration is accomplished by a narrow band filter.

For instance, consider the 30 Mc IF signal in channel A to be A_{if} where

$$A_{if} = F(t) \cos \left[2\pi \mathcal{F}(t)t + \alpha \right] \qquad (3)$$

where F(t) is the amplitude variation with time and $\mathcal{F}(t)$ is the frequency variation with time. Calling the angular frequency of the f_1 oscillator w_1, the oscillator output can be designated by ϕ where

$$\phi = K \cos w_1 t \qquad (4)$$

which is the frequency displacement of channel B with respect to channel A. B_{if} is (as shown in Fig. 4) the following:

$$B_{if} = KF(t + \delta t) \cos\{2\pi [\mathcal{F}(t + \delta t)] [t + \delta t] + a + w_1 t\} \qquad (5)$$

where δt is the time delay or phase shift caused by the angle of arrival at the wave front with respect to the axis of the array. Now, B_{if} is combined with the IF signal from A_{if} yielding an output voltage E_o.

$$E_o = K[F(t)] F(t + \delta t) \cos[-2\pi \mathcal{F}(t)t - a + 2\pi \mathcal{F}(t + \delta t)(t + \delta t) + a + w_1 t] \qquad (6)$$

If the displacement δt is comparatively small, namely, only a relatively small number of rf cycles, we can simplify $2\pi \mathcal{F}(t + \delta t)(t + \delta t)$ to $2\pi \mathcal{F}(t)t + w_1 \delta t$, where w_1 is the carrier frequency of $\mathcal{F}(t)$. This will be practically true if the total frequency excursion for δt is comparatively small with respect to the center frequency, or an equivalent result will be obtained, if the frequency modulation is symmetrical and integration is performed. Simplifying (6) and removing any amplitude variation.

$$E'_o = K' \cos(-2\pi \cancel{\mathcal{F}(t)} t \cancel{- a} + 2\pi \cancel{\mathcal{F}(t)} t + W_1 \delta t + a + W_1 t) \qquad (7)$$

Comparing (7) and (4) in a comparator, $w_1 \delta t$ is obtained (which can be referred to as θ) as a result, which is the angle desired.

It should be pointed out that the receiver will work with any type of signal including noise. All that is required is that the signal have a definite wave front to concentrate the energy into specific phases. This means that there will be no discrimination against any type of signal other than the band selectivity obtained in the RF and 30 Mc IF amplifiers. If there is another unwanted signal in the band of frequencies being received, it will introduce an error which could not be resolved. Random atmospheric noise will introduce noise into the phase comparator and dilute the accuracy.

For good signals, however, such as is obtained in the telemetering band (which is also at frequencies where the atmospheric noise is low), the system has great promise.

Receiver noise will be generated independently at the inputs to each of the RF channels. This means that the noise in one channel will be random with the noise in the other channels and will not correlate. Thus, the noise bandwidth of the system will be the narrowest bandwidth in the chain, which is probably the bandwidth of the phase comparator.

7. Angle Data Determination and Recording

The receiver output is in the form of two sine waves of the frequency f_1 and contains the angle information as a phase displacement with respect to the reference oscillator f_1. In order to extract this data, a phase comparison must be made between the f_1 oscillator and each of the receiver outputs. A direct approach to this problem is to employ a goniometer whose rotor is servo positioned to indicate the desired angle. This, however, limits the data handling capacity to the servo system performance, and precludes using a rapid sampling method for determination of the coarse and intermediate angle data.

A better approach to this problem is to be found in a method which samples the phase data at periodic intervals, and then records the sample value. This would permit rapid switching from one antenna pair to another without an objectionable delay due to the tracking limitations of a servo or similar device.

As is shown in Fig. 5, the output from the f_1 oscillator is fed into a phase shifter which produces a frequency displaced wave whose frequency is the sum of the sweep frequency and f_1. A practical form for producing such a phase shift is to use a phase splitter and goniometer whose rotor is driven at the desired rate by the sweep generator. This rate must be selected so as to provide an adequate number of samples and yet not adversely affect the bandwidth requirements of the receiver. A reasonable rate appears to be about 10 cycles per second. This will provide an adequate number of data points to give a good time history of the trajectory and yet not require an excessive receiver bandwidth.

Using an isolation amplifier to prevent cross-talk, the phase rotating reference is fed into two phase coincidence circuits where a comparison is made with the receiver outputs. When coincidence occurs, a sharp pulse is generated to indicate that a match has been reached. The sweep generator is also used to drive a saw tooth sweep unit where a voltage in linear proportion to the shaft position is generated. When a phase match is reached, this voltage is sampled, smoothed in a filter, then recorded or used in a small analog computer for real time angle determination.

To use the angle data to the highest precision, it must be placed in digital form. This is accomplished by starting a time pulse generator or accurate clock when the sweep generator shaft reaches a reference mark. This starts the counter units so as to count the units of time until coincidence is reached, then the counters are read out and recorded. By this technique

the angle data is placed in digital form, so that the computations for final use can be made to the precision required.

This method of digitalization places a high time stability requirement on the sweep generator. To meet this time stability, the sweep generator is driven from a crystal oscillator in the time pulse generator, thus placing the two on the same time base and eliminating any time drift between them.

8. Coarse-Fine Resolution

To simplify the receiver system, a switching unit is employed to change the RF inputs from the coarse to the intermediate and fine antenna pairs. If a continuous recording could be maintained from the beginning of a flight test, then a resolution of the ambiguities would be required only once. Since this is impractical, a periodic check must be made to assure that the fine data is identified in the correct lobe. This can be made rapidly since the coarse and intermediate angles need be recorded only to the accuracy necessary to resolve the lobe structure from the antenna pair spaced at the next larger interval. This is dependent on the method of phase comparison selected, but can be limited to a minimum of samples necessary for a reliable determination as is shown in Fig. 6. A time sequencer actuates the RF switch and at the same time indicates on the record which group of antennas is in use. Thus the records will show the angles of arrival to the approximate accuracy necessary for ambiguity resolution at periodic intervals and the fine data to the highest degree of accuracy possible during the rest of the recording period.

9. Conclusions

It is concluded that a practical direction finder for use in guided missile trajectory determination is possible through use of cross-correlation techniques, and that such a system is capable of providing data to the accuracy necessary for performance evaluation.

It appears that such a system is not so complex as to make its use at a small down range station impractical, nor is the logistic support required so extensive as to preclude installation on the island chain of the Florida Missile Test Range.

The state of the electronic art has progressed to the point where the circuits necessary to accomplish a system design have been developed. It is, therefore, concluded that no great difficulty will be encountered in undertaking a system development.

APPENDIX I

Phase Difference of Two FM Waves

A frequency modulated wave, E, may be expressed* as

$$E = A \sin (w_1 t + M \sin \Omega t) \qquad (8)$$

where A is the amplitude, w_1 is the carrier angular frequency and M is the modulation index, which in this case is the peak angular phase displacement. Ω is the modulation angular frequency. Delaying this signal by a time δt, the voltage $E_{\delta t}$ is obtained.

$$E_{\delta t} = A \sin \left[w_1 (t + \delta t) + M \sin \Omega (t + \delta t) \right] \qquad (9)$$

or

$$E_{\delta t} = A \sin \left[w_1 t + w_1 \delta t + M \sin \Omega t \cos \Omega \delta t + M \cos \Omega t \sin \Omega \delta t \right] \qquad (10)$$

Taking the angular phase difference, B, between (10) and (8)

$$B = w_1 \delta t + M \sin \Omega t \cos \Omega \delta t + M \cos \Omega t \sin \Omega \delta t - M \sin \Omega t \qquad (11)$$

simplifying

$$B = w_1 \delta t + K \sin (\Omega t + \alpha) \qquad (12)$$

where

$$K = M \sqrt{(\cos \Omega \delta t - 1)^2 + (\sin \Omega \delta t)^2}$$
$$\alpha = \tan^{-1} \frac{\cos \Omega \delta t - 1}{\sin \Omega \delta t} \qquad (13)$$

Integrating (12) to remove the phase modulation, the resultant angular phase difference ϕ is

$$\phi = w_1 \delta t \qquad (14)$$

*Terman, F. E. Radio Engineering, Page 418, Mcgraw Hill Books 1937

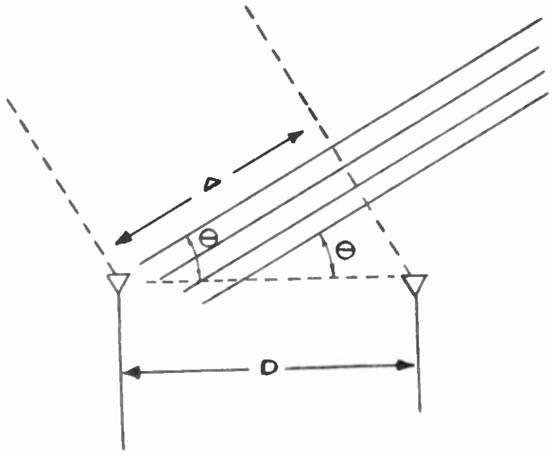


Fig. 1 - Two antennas used to measure direction of wave arrival.

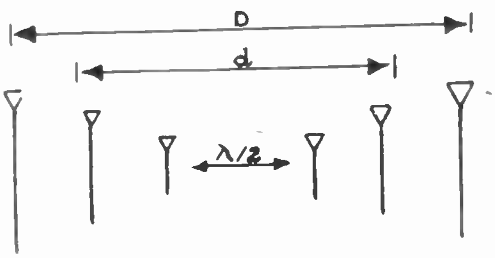


Fig. 2 - Multiple D.F. antennas arrays to resolve ambiguities in readings.

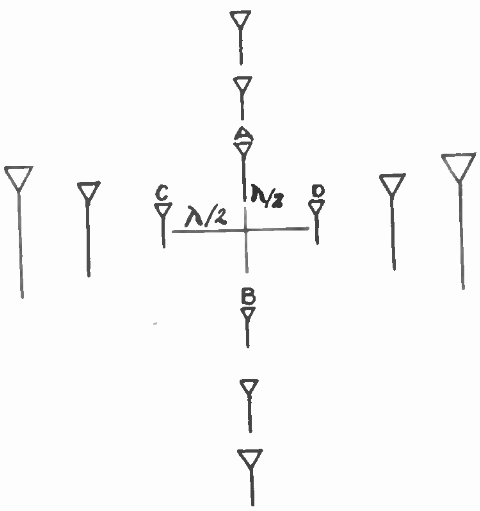


Fig. 3
Plan layout of crossed arrays. D is actually many times larger than d, the size being determined by the accuracy of the phase measuring equipment.

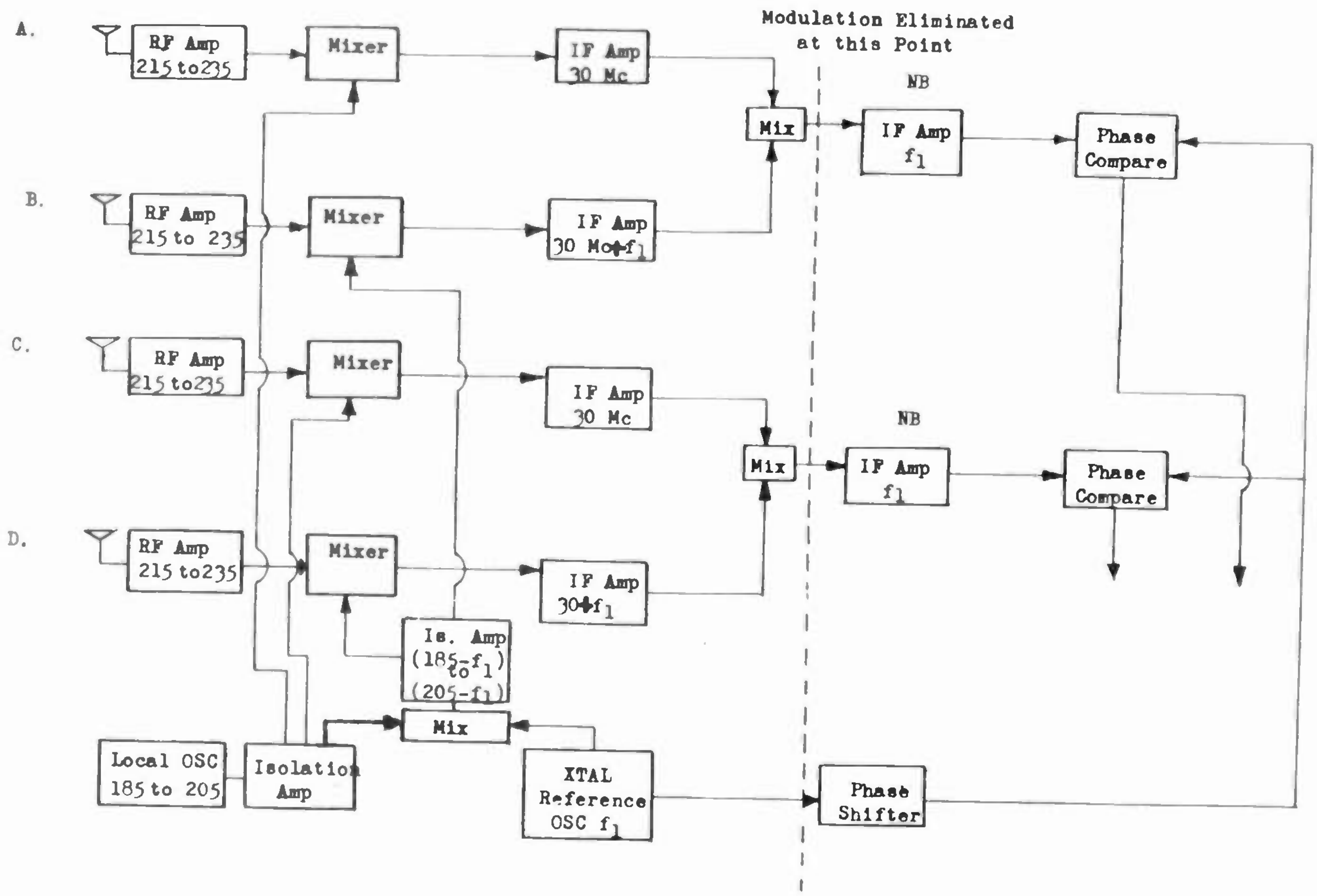


Fig. 4 - Receiver block diagram.

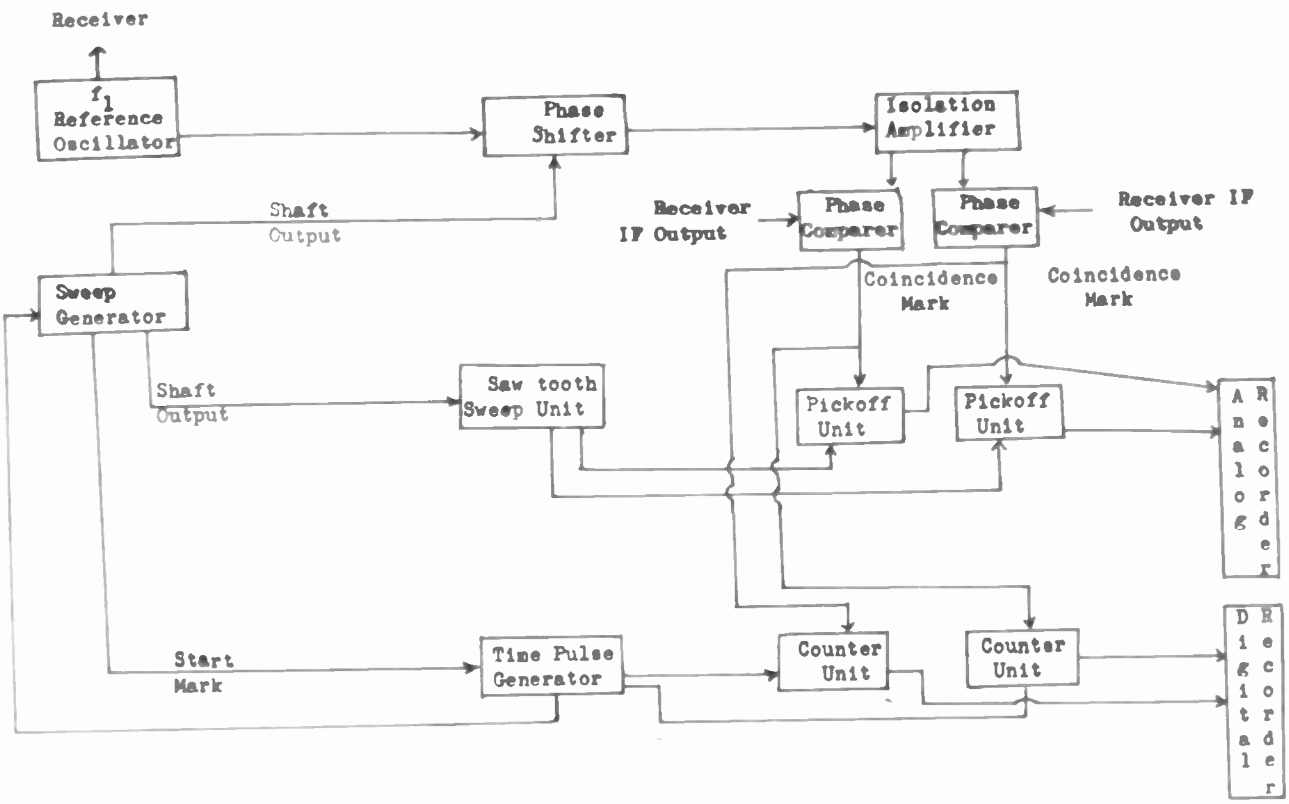


Fig. 5 - Data read-out system.

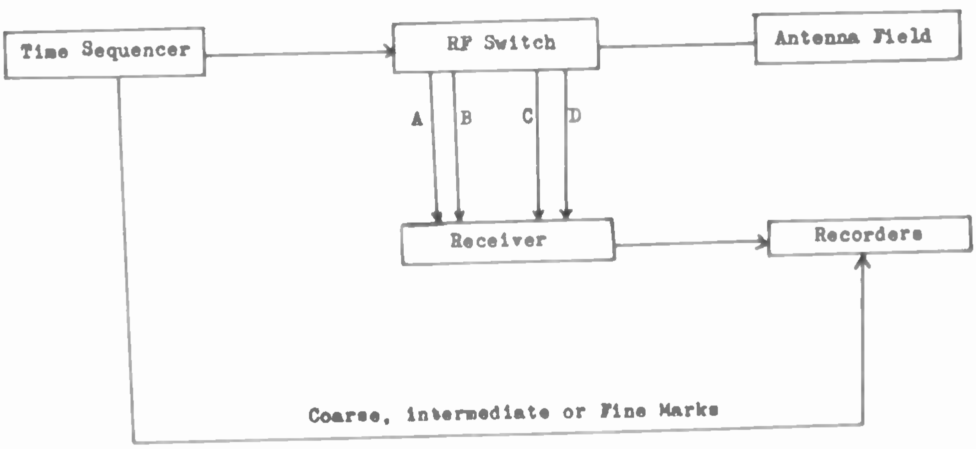


Fig. 6 - Coarse-fine ambiguity resolution unit.

PRESENT STATUS OF MICROWAVE RADIOMETRIC RECEIVER DEVELOPMENT

R. M. Ringoen
Collins Radio Company
Cedar Rapids, Iowa

Abstract

The microwave radiometric receiver was first introduced by R. H. Dicke in 1945. Since that time various methods have been devised for detecting very low-noise powers in the microwave spectrum -- some equipments being capable of detecting noise power changes of less than 10^{-16} watts, which may be equivalent to an antenna temperature change of less than one degree C. The basic theory of microwave radiometry is developed and the various types of receivers used for this work are described and compared.

The introduction of the use of ferrites to obtain unidirectional transmission in waveguide has made it possible to isolate the receiver from the antenna -- resulting in a great improvement in the "switching" type of radiometer at some wavelengths. Waveguide ferrite components adaptable for use in radiometers are described along with the results of development work on ferrite rotators at a wavelength of 8.7 mm.

Thermal Radiation Theory

Radiometric receivers are used mainly in radio telescopes. To understand the problems involved in the design of radiometric receivers one must be somewhat acquainted with the basic aspects of radio astronomy.

Radio astronomy is the science which is concerned with radiation from all celestial bodies in the radio frequency spectrum. Practically all the radiation from celestial bodies in the microwave region is thermal radiation. Any object which is absorbing must also radiate. If a body absorbs all the energy incident upon it it is called a black body. Any body falling into this category radiates according to Planck's law. Planck's law is plotted in figure 1 for various temperatures. From the figure it is shown that the maximum radiation per cycle bandwidth occurs at a wavelength which is a function of temperature -- the higher the temperature the shorter the wavelength. Also the energy radiated at the longer wavelengths is considerably lower. In the microwave region Planck's formula is closely approximated by Rayleigh -- Jeans formula which is:

$$\psi_\lambda = \frac{8\pi k T_s}{\lambda^4}$$

where: k = Boltzmann's constant (1.38 x 10^{-23} joules/mol./deg.k.)
T_s = temperature of black body (degrees Kelvin)
λ = wavelength (centimeters)
ψ_λ = density of radiation -- the radiant energy per unit volume in a stream of radiation at the wavelength, λ

This equation may be used to compute the power received by an antenna when directed toward a black body at a temperature, T. To put the equation in better form both sides are multiplied by $\Delta\lambda$ and the following substitution is made:

$$\Delta\lambda = -\frac{\lambda^2 \Delta f}{c} \qquad \Delta f = \text{detector bandwidth}$$

which yields:

$$\psi_\lambda \Delta\lambda = \frac{-8\pi k T_s \Delta f}{c\lambda^2} \quad (\text{joules/cm}^3)$$

For an isothermal enclosure the following equation holds:

$$\psi_\lambda = \frac{4\pi i_\lambda}{c}$$

where: i_λ = rate at which a body radiates energy in a given direction per unit solid angle and per unit of its own area as projected on a plane perpendicular to the given direction.

Substituting this in the equation for ψ_λ yields:

$$i_\lambda \Delta_\lambda = \frac{-2k\, T_s \Delta f}{\lambda^2} \quad \text{watts/cm}^2\text{/unit solid angle}$$

Hence the power radiated into an incrememtal solid angle will be for one polarization.

$$\Delta P = \frac{K T_s \Delta f}{\lambda^2} \Delta\Omega \quad \text{watts/cm}^2$$

This may be multiplied by the area of the collecting antenna to obtain the total power received by the antenna from the incremental solid angle, $\Delta\Omega$. The effective area of an antenna is given by: $A_e = \left(\frac{\lambda^2}{4\pi}\right) G(\theta,\phi)$

Thus

$$\Delta P_R = \Delta P . A_e = \frac{K\Delta f}{4\pi} G(\theta,\phi)\, T_s(\theta,\phi)\Delta\Omega$$

To obtain the total power received, this equation must be integrated over the total solid angle or

$$P_R = \frac{k\Delta f}{4\pi} \int_{\substack{\text{total}\\ \text{solid angle}}} G(\theta,\phi)\, T_s(\theta,\phi)\, d\Omega$$

If the source has finite boundaries and the background is at absolute zero temperature then the

above expression can be broken into two integrals as below.

$$P_R = \frac{K\Delta f}{4\pi}\left[\int_{Source} + \int_{\substack{Remaining\\Solid\ Angle}}\right]$$

The second integral is zero since T_S for the region is zero. Hence:

$$P_R = \frac{K\Delta f}{4\pi}\int_{Source} G\ (\theta,\phi)\ T_S\ (\theta,\phi)\ d\Omega$$

When the antenna is directed towards a source of constant temperature, an effective antenna temperature T_R may be defined such that: $P_R = KT_R\Delta f = \frac{KT_S\Delta f}{4\pi}\int_{Source} G\ (\theta,\phi)\ d\Omega$ or: $\frac{T_R}{T_S} = \frac{1}{4\pi}\int_{Source} G\ (\theta,\phi)\ d\Omega$

This relationship is called the antenna factor and indicates the amount the antenna temperature is reduced below the source temperature.

There are three cases to be considered. If the antenna is completely surrounded by a source at temperature T_S, the integral is equal to 4π and $P_R = K\ T_S\Delta f$. This same relationship holds true if the source is of restricted size, but with the antenna beam completely confined to the source. Increasing the gain of the antenna beyond this point only permits radiation from specific areas of the source to be studied.

The equation $P_R = KT\Delta f$ reveals the amount of received power on which a radiometric receiver must operate. Suppose that it is desired to detect a one degree C. change in antenna temperature and the receiver bandwidth is 16 megacycles. The change in power will be

$$P = KT\Delta f = 1.38 \times 10^{-23} \times 16 \times 10^{6}$$

$$P = 2.21 \times 10^{-16} \text{ watts}$$

If the receiver has a noise figure of 14 db (25X) the noise output will be equivalent to an antenna temperature of 7500 degrees K. A one degree antenna temperature change would then produce a signal—to—noise ratio of 1/7500 or 0.013%. This illustrates clearly that if such an antenna temperature change is to be detectable special procedures must be used in the design and construction of the receiver.

Three basic types of receivers have been developed for use in radio telescopes. Each of these three basic receiver types will be discussed in general.

D. C. Type Radiometer

The first to be considered is the "d-c" type. The block diagram of this receiver is shown in Figure 2. The noise power received by the antenna is amplified in either a tuned radio frequency or a superheterodyne receiver. The amplified noise is then detected in a diode detector. The output of the detector is amplified in a d-c amplifier and recorded on a recording meter. Although in block diagram form the receiver appears to be simple in actuality it is quite complex. It is made complex by the very rigid gain stabilization which must be realized in the receiver. If the receiver noise figure is 20X the output of the receiver when its antenna is terminated in a matched load at room temperature is equivalent to a 6000 degree antenna temperature. If the receiver is to detect a few degrees change in antenna temperature the receiver gain must be about 120 db. A one percent change in this gain or in the noise figure would then cause an output change equivalent to a 60 degree antenna temperature change. This indicates the need for extremely good gain and noise figure stabilization in the receiver. This generally resolves itself into multistage voltage regulators for all plate and filament voltages in the receiver and temperature stabilization of the complete receiver enclosure. In practice the receiver will be seldom turned off as the time required for the gain to stabilize is considerable.

Since the output of the diode detector will be proportional to several thousand degrees this voltage must be nullified if small temperature changes are to be measured. This is done by a d-c voltage from a very stable d-c source. The difference between the detector output and nullifying voltages is then amplified in a d-c amplifier and presented on the meter. The d-c amplifier also presents a problem in gain stabilization.

Receivers of this type have been built to detect antenna temperature changes of a few degrees in the VHF and UHF regions. However, use of this type of receiver has not been too extensive in the microwave region because of the difficulty in stabilizing the gain and noise figure of the crystal mixers required.

Servoed Noise Source Radiometer

The second type of radiometric receiver to be considered is the servoed noise source type. The block diagram of this general type of radiometer is shown in Figure 3. In this receiver a motor driven switch alternately connects the receiver input to the antenna and to the output of a variable noise source. The signal is amplified in a tuned radio frequency or superheterodyne receiver. The detected output is applied to a synchronous demodulator as is the output of an a-c generator driven at the same rate as the r-f switch. The d-c output of the demodulator is then proportional to the difference between the equivalent antenna temperature and the temperature of the noise source. This voltage is in turn used to vary the noise output of the variable noise source. If the noise source is a temperature limited diode the demodulator output would act to control the diode anode current so that the noise output would be equal to that received by the antenna. In this manner the diode current would be an indication of the antenna temperature, and would be so calibrated and recorded on a meter.

In this type of radiometric receiver the receiver gain may vary without affecting the output

directly, as may the receiver noise figure. Also the receiver need not be calibrated because the noise source calibration is the receiver calibration and this is not affected by receiver gain variations. Receiver and detector linearity are also of secondary importance.

Since a matched noise source cannot be equivalent to a termination below ambient temperature this servoed noise source type of radiometer can be used only for measuring temperatures above ambient. This is a disadvantage. Also, an easily controlled and calibrated noise source must be available. At present this fact limits the use of the receiver to the UHF and VHF spectrum. Finally, since some small difference in antenna and noise source temperature must be present to supply the noise source control element some error will exist. Consequently, the sensitivity of this type of radiometer, not considering receiver gain changes, is somewhat poorer than the d-c type or the Dicke type which will be discussed next.

Dicke Radiometer

The block diagram of the general Dicke-type radiometer is shown in figure 4. In this type of radiometer the receiver input is alternately terminated by the antenna and an eccentive absorbing disk which dips into the antenna line waveguide. The eccentric absorbing disk is driven by a synchronous motor which also drives an a-c generator. Consequently the noise power applied to the input of the receiver is modulated at a frequency equal to the motor speed and the peak-to-peak modulation is proportional to the difference in temperature between the antenna and the reference or modulator wheel.

A balanced r-f mixer is generally used in the Dicke-type microwave radiometer. This is done to minimize local oscillator noise and to reduce to a minimum the local oscillator power coupled into the antenna line. The two outputs of the balanced mixer are coupled into an i-f amplifier through a balanced - to - unbalanced transformer. The i-f amplifier will generally have about 100 db gain. The output of the diode second detector is amplified in an audio amplifier which is tuned to the modulation frequency, usually 30 cps. The output of the tuned audio amplifier is then demodulated in a phase detector or synchronous demodulator. The synchronizing or reference voltage is supplied by the a-c generator driven by the synchronous motor which drives the reference wheel.

The d-c output of the demodulator is integrated in a low pass filter and recorded. This output is proportional to the difference in temperature between the antenna and modulator wheel. The bandwidth of the detector may be made as small as desired -- effectively narrowing the band about the exact modulation frequency. This results in a very sensitive receiver. Receivers of this type have been built which are able to detect an antenna temperature change of 0.1 degrees centigrade — equivalent to an input power change of 2×10^{-17} watts and a signal-to-noise ratio at the second detector of one part in 25,000.

The Dicke-type radiometer is calibrated by terminating the antenna in a matched load, the temperature of which is known. The temperature of the modulator wheel of course must also be known.

Comparison of Radiometer Types

In table 1 a comparison of the three types of radiometers is made with some of the advantages, disadvantages and major uses of each listed. The d-c type of radiometer since it has no modulating frequency may have as large a detector bandwidth as the bandwidth of the output recording device. As such it is useful in studying the fine grain structure of the radiation from various sources in the VHF and UHF regions. However, the fact that its gain and noise figure must not vary as much as 0.1% during the measurement period is a definite disadvantage.

The servoed noise source type of radiometer was conceived to eliminate the effects of receiver gain variation and non-linearity and to simplify calibrating the receiver in terms of antenna temperature. Doing this requires a stable noise source, the output of which may be calibrated and continuously varied over a wide range. Switching from the antenna to the noise source also causes undesirable transients in the output for microwave receivers. Since the effective temperature of the noise source cannot go below its actual temperature if it is matched the servoed noise source type of radiometer may not be used for measuring very low temperatures.

The Dicke-type radiometer is the simplest in that it requires no extreme gain or noise figure stabilization, no r-f switching and no high temperature noise source. In general it is the most sensitive and versatile type. It's main disadvantage is that it requires a modulator wheel which is well matched for all positions, making the wheel's size become large for the longer wavelengths. The Dicke-type has found by far the greatest use in the microwave region and is the type which we have built at Collins Radio Company.

Use of Ferrite Waveguide Components

The extreme sensitivity of the radiometer requires the consideration of some factors not normally considered in receiver design. Local oscillator and crystal noise in the image and signal bands can cause an output in the radiometer if coupled into the antenna line and then reflected back into the receiver input. The power level may be 10^{-14} watts or less but this can cause an output proportional to 50 degrees antenna temperature. This residual output will vary with frequency and is generally the limiting factor in the accuracy of measurement when the Dicke radiometer is used to measure small temperature changes or low absolute temperatures. This effect may be eliminated if no power from the r-f mixer is coupled into the antenna line and permitted to be reflected back into the input--appearing as signal.

The introduction of the use of ferrites in waveguide and the utilization of the Faraday rotations obtainable has resulted in the development of components most suitable for use in microwave radiometers. In Figure 5 is shown a cylindrical waveguide containing an axially mounted ferrite rod surrounded by a solenoid. When a current is passed through the solenoid, the resulting magnetic field in the ferrite causes the polarization of the wave in the cylindrical guide to be rotated. The wave will be rotated in the same direction for both directions of transmission. This non-reciprocal characteristic allows several interesting waveguide components to be built. One of these is a unidirectional transmission line. The uniline is made up of a section of cylindrical guide containing a ferrite rod and an axial magnetic field such that the polarization rotation is 45 degrees. The input and output guides are rotated 45 degrees with respect to each other. For one direction of transmission the polarization rotation through the ferrite is such that the emerging wave is of the same polarization as the output rectangular guide so that negligible loss is encountered. For the other direction of propagation the 45 degree rotation produced by the ferrite causes the polarization of the emerging wave to be 90° from that of the output waveguide so that little transmission is possible in this direction. (A resistive wafer is generally inserted to absorb the energy of the wrong polarization). Hence, a unidirectional transmission line may be built with about 0.2 db loss in the forward direction and 20 to 30 db loss in the backward direction.

Placing a unidirectional transmission line between the r-f mixer and modulator wheel so that power may flow only towards the mixer has been found by Mr. C. H. Mayer of NRL to reduce by 90% the residual output of an X-band Dicke radiometer. The addition of this uniline thus permitted the radiometer to be capable of measuring absolute temperatures with an accuracy of a few degrees centigrade.

Another ferrite component of interest is the microwave gyrator shown in Figure 5. In this component the magnetic field is adjusted to yield a 90 degree polarization rotation through the ferrite. For one direction of transmission this rotation adds to the 90° rotation in the waveguide twist and in the other direction the two subtract. Thus, the gyrator may be considered to have 180 degrees phase shift for one direction of transmission and zero for the other direction. Reversing the current flow through the solenoid reverses the phase shift relationship. This property allows the gyrator to be used to effect a unique waveguide switch which may be used to replace the modulator wheel in a Dicke radiometer and to prevent any power emerging from the r-f mixer antenna line from being modulated and reflected back into the mixer input. The switch is called a microwave circulator and is shown in figure 6. The gyrator is placed in a line connecting two colinear arms of the two hybrid junctions and the other two colinear arms are joined by a line of length equal to that of the line containing the gyrator for one direction of current flow through the gyrator solenoid.

Inspection of the circulator reveals that for one direction of current flow through the solenoid energy from the antenna is coupled into the receiver input line and for the other direction of current flow the reference load is connected to the input. For both directions of solenoid current flow power emerging from the receiver input when partially reflected by the antenna or reference load ends up being dissipated in the matched load -- thus reducing greatly any residual receiver output. Mr. Mayer of NRL has found that use of a circulator at X-band reduces the residual receiver output to 2% of that measured for a Dicke radiometer using a modulator wheel. This makes the radiometer excellent for measuring small absolute temperatures accurately. Also the reference temperature is the temperature of a waveguide load, not an absorbing wheel, and may be more easily stabilized or varied for receiver calibration purposes.

At Collins Radio Company we have been engaged in the development of microwave gyrators at the wavelengths of 1.9 cm and 8.7 mm. Since no work at these wavelengths has been previously reported our results to date may be of interest. In Figure 7 is shown the rotation vs. magnetic field for one ferrite rod sample mounted in circular guide at a wavelength of 8.7 mm. A field of 43 oersteds (20 ma. current through the solenoid) produced a rotation of 90 degrees. The VSWR looking into the section remained below 1.05 for all magnetic fields up to 300 oersted and the loss was only 0.3 db. This compares favorably with results obtained at longer wavelengths. We have obtained similar results at 1.91 cm -- indicating that ferrite components are effective at both these wavelengths.

The inception of the uniline and microwave circulator have added considerably to the quality of measurement which can be made with a radiometer. Use of a uniline in a Dicke radiometer between the modulator and r-f mixer or use of a circulator to replace the modulator wheel in a Dicke radiometer makes this type more conclusively the most sensitive, most easily operated and most versatile of all radiometer types.

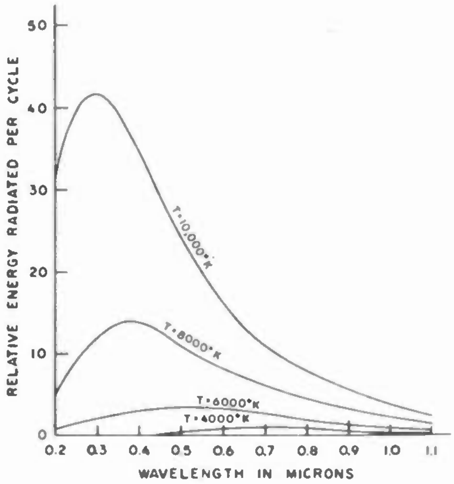


Fig. 1 - Plot of Planck's radiation law.

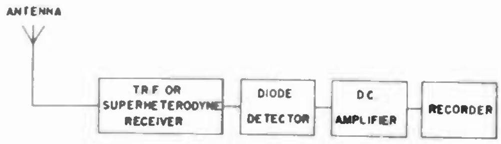


Fig. 2 - DC type radiometer.

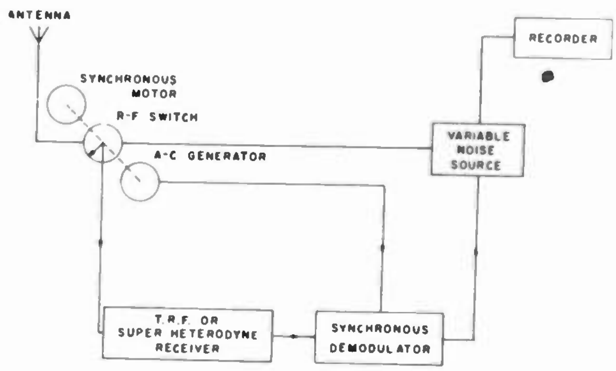


Fig. 3 - Servoed noise source radiometer.

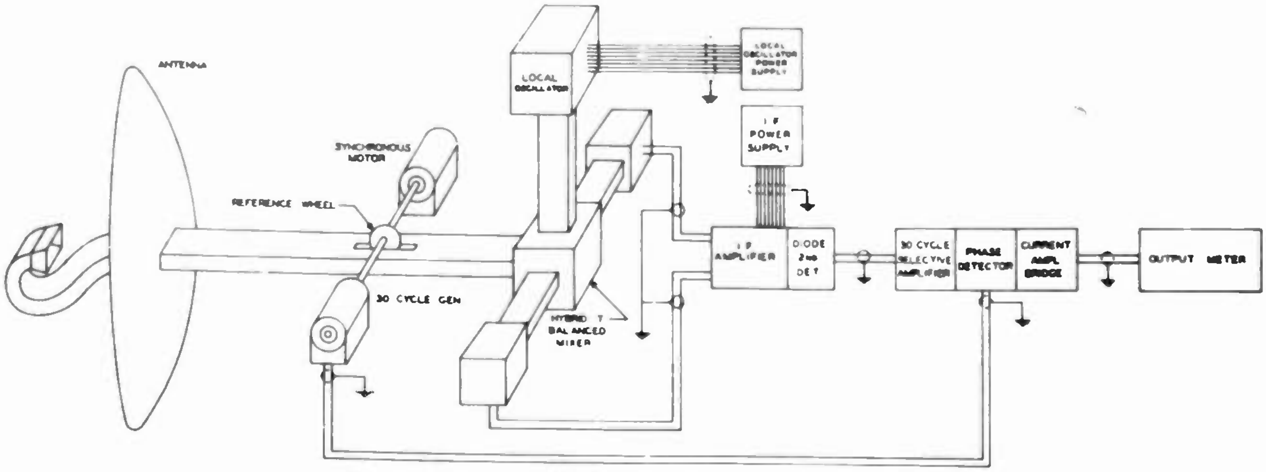


Fig. 4 - Block diagram of general radiometer.

COMPARISON OF THE THREE BASIC TYPES OF RADIOMETERS

	D. C.	Servoed Noise Source	Dicke, Reference Wheel Type
Advantages	Large Detector Bandwidth No switching transients	No gain stabilization needed No receiver calibration required Detector Non-linearity eliminated	No extreme gain stabilization required Minimum system complexity Maximum system versatility Maximum sensitivity
Disadvantages	Extreme gain stabilization required Receiver calibration required Noise figure must not vary	Cannot measure low absolute temperatures Requires difficult r-f switching Requires stable, high output noise source Poor sensitivity	Receiver calibration required Reference wheel becomes large at longer wavelengths
Major Use	Non-portable VHF and UHF Radiometers for measuring large temperatures	UHF Radiometers for measuring temperatures above ambient	All microwave receivers below 10 cm wavelength where good long term stability is required

Table I

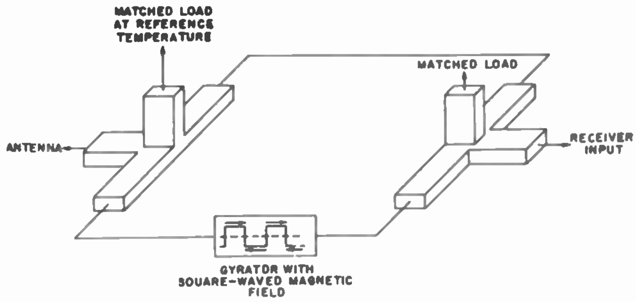


Fig. 6
Microwave circulator as used in Dicke-type radiometer.

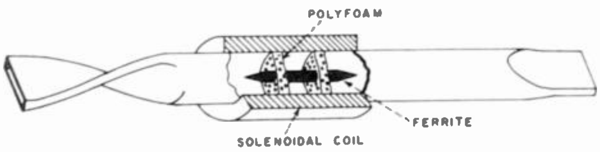


Fig. 5 - Microwave gyrator.

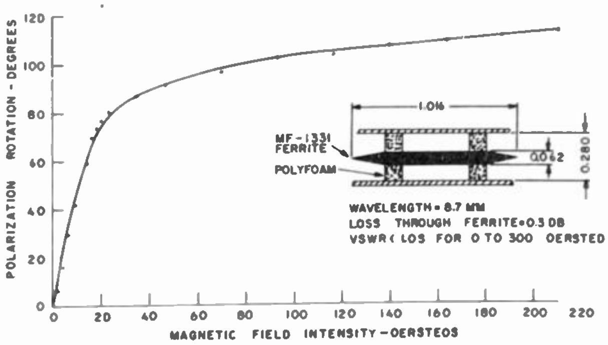


Fig. 7
Polarization rotation vs magnetic field for cylindrical waveguide containing axially mounted ferrite rod.

GUIDED MISSILE RANGE INSTRUMENTATION
A NEW ELECTRONIC ART

M. S. Friedland
Air Force Missile Test Center
Patrick Air Force Base, Florida

INTRODUCTION:

Guided missile warfare is not a new concept for the Germans were developing, experimentally, air and surface launched missiles long before World War II. However, the full import of the guided missile as tactical weapon was brought forth in the latter phase of the last World War with the familiar "Buzz-bomb" attacks.

Allied intelligence revealed that in 1943 the congealed efforts of German engineers were consolidated into 12 weapons to be carried through to immediate development for operational use. Further, the dissemination of technical information in the post-war period revealed that the progress at Peenemunda had reached menacing proportions. With a short extension of time, the Germans could have produced a surface to surface missile which may have made the V-2 appear as a crude experiment and the outcome of that war might then have had an entirely different aspect.

The potentialities, then, of these weapons immediately became apparent. A profound study of the problem revealed one very obvious fact: if the United States is to keep abreast of the world in military strength and weapons it must combine its resources into the development of long range guided missiles. It must also provide the facilities to accommodate testing many long range pilot-less weapons.

Consequently, the first session of the 81st Congress enacted a law for the establishment of such a test facility; Public Law 60 which authorized the Secretary of the Air Force to establish a "Joint Long Range Proving Ground for Guided Missiles and Other Weapons". The Chief of Staff, USAF, in Air Force Regulation 20-65 defined the mission of the Range as follows:

> "The Long Range Proving Ground Division will establish, operate, and maintain a long range designed for the purpose of meeting the requirements for pre-flight and flight testing in connection with component and vehicle development, operational evaluation, and training in weapons systems which require such a range."

A search of the United States for a suitable geographic location for the Long Range Missile Range led, finally, to the selection of the present one. It was the unanimous opinion of a joint committee that no other site, with its launching area within the continental limits of the U. S., could be better suited to the purpose of the Long Range Proving Ground. From this site on the eastern coast of Florida, the flight line would be directed in a southeasterly direction to Puerto Rico permitting installation of land based range instrumentation sites in the Bahama Islands. Secondly, the equitable climate permits year round operation. The area is also relatively uninhabited and, finally, the decommissioned Banana River Naval Air Station could be reactivated, accommodating the Long Range Proving Ground Division.

In 1948 the Navy transferred to the Air Force the Naval Air Station at Banana River, and the initial phase of the development of the range began. The Commanding General, LRPGD, was directed responsibility for initiating action to secure foreign agreements for rights and access to land and/or waters under foreign jurisdiction required by the LRPGD and activate the range. The scope of the tests anticipated, operational considerations, and limitations of the electronic art led to the division of the present range into nine operating Range Stations located in order at Cape Canaveral, Jupiter Inlet, Grand Bahama Island, Eleuthera Island, San Salvador, Mayaguana Island, Dominican Republic, and Puerto Rico as shown in Figure 1. The mission of the Long Range Proving Ground established the task of development, modification and installation of electronic, optic, and communication equipment so as to fit out these stations and provide an integrated Guided Missile Range instrumentation system.

RANGE INSTRUMENTATION:

Range instrumentation is defined as that equipment which is necessary to measure and evaluate the performance characteristics of guided missiles. Broadly speaking, the composite instrumentation to be used throughout the range is comprised of trajectory equipment,

telemetering equipment, data processing equipment, and the requisite supporting facilities.

Two basic complementary methods of measuring missile performance are available at present. The first depends on external observation of missile trajectory which can be determined either from electronic or optical measurements. The second method depends on measurements of internal missile characteristics, which are remotely recorded by radio telemetering those characteristics as required by a missile developing agency. The basic systems just described must be augmented by all necessary supporting facilities such as communications, data relaying, range safety, and range clearance.

EXTERNAL INSTRUMENTATION:

External instrumentation is composed of all equipment which is used to determine missile space trajectory. The equipments in use on the range determine the space position, velocity, acceleration, and attitude of the missile with varying degrees of precision. Under the present program rough trajectory information is obtained from the range chain radar system for use in real time, and precise trajectory data is obtained from optical and electronic base-line systems as well as single station precision radars. The collection of acceleration data is usually restricted to telemetering instrumentation and external optical determinations during the launching phase.

ROUGH ELECTRONIC TRAJECTORY SYSTEM:

Study of the criteria of missile trajectory instrumentation resulted in the selection of an S band auto-tracking chain radar system for rough electronic measurements and range safety. The principle components of the chain consist of an analog computer, electronic plotting boards, recorders, and data transmission equipment in addition to the basic radars. Figures 2, 3, and 4 illustrate typical installations of these components.

The principle of the chain radar system depends on the radars being located at strategic points along or near the flight line to provide automatic tracking. Test missiles equipped with radar beacons respond to the signals of each radar station within its range. The missile position data is then transmitted to the next station for acquisition and tracking as the missile progresses along the range, and to Central Control at Cape Canaveral. Space position information of the missile is automatically plotted at each station for information and judgment of the interested agencies. Aberrations in flight path are readily observed, and action is taken accordingly.

PRECISION ELECTRONIC TRAJECTORY SYSTEM:

The accurate analysis of missile performance as required by missile contractors led to the investigation of all presently known electronic space position systems. These systems can be divided into two general categories:

(1) Single station systems whose operation requires measurement of range and angles.

(2) Triangulation Systems whose operation depends upon measurements of ranges from fixed points on long base lines.

Errors resulting from radio-propagation variations affect the accuracies of measurement of low angles in the single station systems and, consequently, these will be augmented by a series of base-line triangulation systems.

Currently under test is a three dimensional Raydist system to provide base-line measurements for the interim period. A signal from a CW transmitter in the missile is received together with a second signal from a ground based reference transmitter at four stations on the ground. The signals are locally phase compared, and the phase difference signals are relayed to the Master Station at or near one of the receiving sites. At the Master Station selected pairs of these signals are again phase compared and the resultant Signals are a measure of the location of the transmitter with respect to three given base-lines. The accuracy of the system degenerates as the missile distance from the base-line increases. In addition, a precision radar and a direction finder system taking bearings on the telemetering signal are under development to meet the precision requirements.

PRECISION OPTICAL TRAJECTORY MEASUREMENTS:

Optical instrumentation is used to provide missile trajectory, attitude and time and motion study data at the launching area and impact areas. In the launching area, a group of phototheodolites, in conjunction with long focal length tracking telescopes are used for position and attitude determination. In addition, precise trajectory and pitch information of the initial portion of trajectory is recorded using a series of acceleration cameras. Augmenting

this, Fastex 16mm and 35mm cameras as well as high speed Mitchell Chronographs all modified for electronic control and timing are positioned as required for time and motion studies. Figure 5 is an illustration of a typical photo-theodolite system, Figure 6 shows a tracking telescope, Figure 7 illustrates an acceleration camera with its associated equipment.

Impact area position and attitude information is likewise derived from photo-theodolites and tracking telescopes. Target acquisition information for the optical instruments is provided by the range radar chain. Data transmission equipments similar to those provided in the chain radar system are used to relay acquisition information to electronic parallax computers associated with each optical site.

INTERNAL INSTRUMENTATION:

Internal instrumentation includes techniques and facilities which are used for measuring the performance of components of the missile. Radio telemetering techniques provide the principal methods for remotely measuring these quantities; however, supplementary missile performance data may also be obtained from internal recoverable recording devices.

Telemetering systems as presently envisioned can be divided into three general categories; FM/FM systems, Pulse Width Modulation systems, and Pulse Code Modulation systems. Standardized FM/FM and Pulse Width Telemetering equipment has been selected for use on this range. Pulse Code systems are being investigated for future requirements.

The purpose of this telemetering system is to record at a group of ground receiving stations the information originating from end instruments in the missiles while in flight. With the standard FM/FM system, telemetered data is received and presented as a voltage or current output with an accuracy of about 1% of full scale. Figure 8 shows a typical telemetering receiving antenna system and Figure 9 is of a FM/FM ground station.

To simplify the collection, transmission, and processing of telemetered data throughout the range, data is collected from down range stations on magnetic tape or transmitted to Cape Canaveral via the Submarine Cable System. Several discriminator racks, capable of generating sixteen telemetering sub-carrier channels and decomutation of these as required, are available at Cape Canaveral for real time and post flight processing of all telemetered data.

Prior to the full operation of the submarine cable similar such racks, including the associated recording equipment are used at each down range station for real time requirements and permanent records.

DATA REDUCTION AND ANALYSIS:

All pertinent external and internal missile performance data which is collected during a missile firing is so processed as to permit a detailed evaluation of the missile performance with respect to time of flight. The range instrumentation system records, raw data on motion film, fixed photographic plates, magnetic tape, etc must be edited, calibrated, correlated, and processed before it can be assembled in report form as tables and curves to describe a missile flight. This processing is performed at Patrick AFB.

Tracking telescope information is analyzed by simple analog techniques, wheras processing of the other missile trajectory data must be performed by a large scale electronic digital computer developed at the center. Since most raw trajectory data is not suitable for direct introduction into the computer, a data preparation phase is required before final computation. Precision radar and base-line triangulation data, as well as rough trajectory radar data, can be fed with simple coding to the computer. However, the photographic records from precision optical instruments and the interim Raydist system must be manually edited and read before being processed. Since internal performance data can seldom be recovered with an overall accuracy of better than 1%, processing of such data is accomplished without appreciable deterioration by means of analog linearizers. Therefore, telemetered internal performance data will seldom be processed by the digital computer, but may be digitized to permit tabulation when required. Figure 10 is an illustration of the large scale digital computer developed at the center and known as FLAC.

SUPPORTING FACILITIES:

It is obvious that in addition to the basic measurement and data reduction systems which have been described, certain supporting facilities are required in order to properly integrate the operations of the range.

RANGE CLEARANCE:

Range Clearance is the Center's responsibility for surveillance of the range and clearance of surface vessels and aircraft to reduce

to an acceptable probability the danger resulting from a missile test. Aircraft and shipping must be cleared from the range prior to missile launching to permit unrestricted use of the proposed flight path. A system of surveillance radar networks so situated as to cover the entire range supplies information to investigator aircraft and to ships for direction of range clearance operations. To enable the range clearance operations to be conducted over desired portions of the range, Range Clearance Centers are established at Cape Canaveral and at all major down range stations. Range Clearance Centers contain radar and communications facilities necessary to receive, display, and disseminate information concerning present and predicted range conditions and for the tactical control of aircraft and surface vessels assisting in the range clearance operations.

Range surveillance is divided into two categories: an airborne radar system for surface coverage, and a surface based radars for air surveillance.

The airborne system employs radar units in multi-engined aircraft stationed over the major shipping lanes throughout the range with associated repeater stations located on appropriate range stations. The primary radar information is coded into pulse form before relaying to the ground installation. Data received at each ground station is plotted manually on a surface plot covering that particular station's area of responsibility. When possible, information on the plot will indicate the type of vessel, course, speed, and time. Secondary information for the plot may also be obtained from investigator aircraft and crash boats.

The ground based air surveillance radars provide a display of the area surrounding each Range Clearance Center. Contact reports are made directly from the main radar consoles, and plotted on a vertical plotting board, thus maintaining a current status plot of all aircraft in the danger area.

In addition to the above mentioned operations, which are common to all Range Clearance Centers, the Cape Canaveral Center maintains a plot of the entire range used for the flight test. Information for plotting on this board is received from all stations via radio or submarine cable. Figure 11 shows a typical Range Clearance Center.

From the information gathered by the radar network and displayed on the various plotting boards, the Range Clearance Staff determines the position of all questionable traffic and dispatches the Investigator aircraft and vessels as necessary to clear the range. The investigator aircraft and vessels serve as messengers to pass warning information to traffic within the danger area by means of various communication equipments. Permanent records of all operations are made at each clearance center by name of magnetic voice recording and photographic techniques.

FLIGHT SAFETY:

The range flight safety system is based upon an analysis of the missile trajectory plotted in real time. For this purpose the real time trajectory is determined from the range radar system, as previously described under "Rough Electronic Trajectory System." The charts which the trajectory plot is presented include destruct lines and danger areas. This data enables the Flight Safety Officer to make a decision as to a safe or unsafe flight and whether or not to destroy the missile.

The destructor system consists of a specially developed radio control transmitter, and a receiver in the missile. The equipment uses a tone modulation to provide twenty channels of on-off control on one RF carrier. Both command guidance and destruct functions are accomplished by this radio equipment. Destruct functions utilize four of the channels, while the remainder are available to the using agency for guidance purposes.

For destruct purposes, the four channels are used to actuate relays in the receiver so connected as to cut off the fuel supply or fire explosive squibs to destroy the missile when commanded from the ground. A modified "fail-safe" method of missile destruction with command override is employed by the range. By this method carrier signals from the destruct transmitter are interlocked with the radar beacon signals to provide automatic destruction when neither the radar nor the destruct transmitter signals are received at the missile.

Two destruct transmitters, one a standby, are located at Cape Canaveral, and at each major down range station.

TIMING SYSTEM:

The range timing system supplies an accurate time reference for all measurements made throughout the missile range as well as the basic control rate for all automatic firing equipments. The space-position information obtained from the optical and electronic tracking systems is thus correlated in time, making it possible to compute space-position acceleration, and velocity at any desired instant.

The timing system as developed for the range is designed to provide starting time, elapsed time, and synchronizing pulses to all instrumentation located in the launching area, at down range stations and at all designated impact areas. The Cape Canaveral time-standard equipment is located at central control and generates all timing and control signals for local use, together with a binary-coded identification of each time interval. All timing and control signals, in the Canaveral area, are distributed by underground cable where this method of transmission is feasible. The outlying theodolite stations receive their timing and synchronizing signals via an FM Radio Link.

Coded tone bursts including identification at the rate of one per second generated by the time-standard and transmitted to all down range stations via a 3kc submarine cable channel for synchronization purposes. Due to transmission time delays between Cape Canaveral and the down range stations, it is necessary to advance the cable tone bursts a sufficient amount to permit the installation of variable time-delay equipment at each down range station. As an interim measure until the submarine cable is completely operational, a HF radio link is being used.

Secondary timing generators are installed at each down range station, and are synchronized with the timing standard by the signals from the submarine cable. Outputs of these secondary timing generators plus binary coded time identifications are distributed to the various instrumentation sites in each area either by wire or VHF radio circuits.

The time base generators are each comprised of a stable crystal oscillator, a number of electronic divider circuits, pulse shaping circuits, cathode followers, and d-c amplifiers- - all rack mounted as a single assembly. The down range time base generators differ from the main base generator only in complexity. Figure 12 illustrates a timing generator for a down range station. Inasmuch as coded time indications are transmitted down range from Cape Canaveral, coding equipment is not required down range except at impact areas where photo-theodolites are to be used since the rate must be converted from the code pulses being transmitted down the cable to a new rate suitable for this application.

Timing for Ribbon Frame Cameras require 1000, 100 and 1 per second pulses with time identification for each second. Bowen CZR-1 cameras will require an additional 50 kc timing signal which will be generated by terminal equipment at the camera site.

Timing for the range radar recorders requires one-per-second pulses beginning at Reference time.

Timing for photo-theodolites requires control signals at the rate of 4, 2, or 1 per second with binary code identification of each second. Equipment is installed at each theodolite to provide a variable time adjustment between shutter actuation and flashlamps to illuminate the data scales so that the flashlamp may be operated at full shutter opening.

Timing signals for the telemetering records is required at the rate of one-per-second with binary code identification. An FM modulator unit is furnished to convert this d-c information into an FM signal at 40 kc for insertion into the magnetic tape records.

MISSILE FIRING SYSTEM:

In the launching of a missile, properly tied sequencing operations are necessary to fire the missile, to provide reference time indexing, and to actuate the necessary range instrumentation equipment. Since the missile is the prime responsibility of the testing agency, the sequencing equipment for firing the missile will be furnished by that agency. However, instrumentation equipment is the responsibility of the range, and for this reason the sequencing equipment which provides reference time indexing, time of first motion, and actuation of the instrumentation equipment must be furnished by AFMTC.

Synchronization of the testing agency's automatic sequencer with the base automatic sequencer is mandatory to permit the base sequencer to furnish necessary negative time signals. This synchronization will be accomplished at the time at which the contractor initiates his sequence by delaying his initiating pulse until the next second mark is received from the timing system.

Provision is made to determine first motion of the missile. Insofar as firing is concerned, AFMTC must necessarily have control of the firing circuit prior to launching. To this end, the Test Control and Launching Area Officers are provided with certain safety controls as follows.

The launching Area Officer and the Test Control Officer each has at his disposal switches which are located in series with the main firing lines to prevent missile firing in case an unsafe

condition arises in their area of responsibility. The Test Control Officers console is shown in Figure 13. In addition, safety plugs are inserted in all firing lines, both at the launching pad and in the blockhouse, by the person in charge when the area is clear. However, before insertion of the safety plugs, voltmeters are available to be used to check the firing lines at both plug stations, and a continuity meter will be used by the Launching Officer to permit a last minute check of all firing lines and ignitors.

Two blockhouses have been constructed to date for general use at Cape Canaveral. One firing system is installed in each blockhouse, and will be used with missiles on either of two launching pads associated with each blockhouse. Six firing lines from each blockhouse to its two subsidiary launching pads (24 lines) are provided for the necessary control. All switching of controls, indicators, and communication lines is accomplished in the communication distribution station.

COMMUNICATIONS:

The range communication system is designed to provide complete facilities for the accurate and expeditious interchange of all information needed on the missile range. Land lines, submarine cable, and radio circuits are combined into an integrated system interconnected through switchboards and patching panels to provide the required flexibility. In addition to carrying voice and telegraph traffic, the communication system provides for such functions as the remote control of equipment, operation of alarm systems, distribution of timing signals, and dissemination of data. The system may be divided into two broad categories: administrative and technical.

The administrative portion of the system provides the telephone and telegraph service which is required to conduct the day-to-day business of the range. It may be considered as the basic system, supplementary to, but separate from, the more complex system required during missile test operations.

The technical portion of the system provides the additional and exacting communications service required just prior to and during a missile operation. Since much data must be gathered and simultaneously distributed during this period, networks are formed and assigned for the exclusive use of the several operational groups. For example, the groups concerned with Range Clearance and Safety required networks capable of providing instant communication between personnel located on land stations, aircraft, and on ships. Groups primarily concerned with measuring and recording flight test data require additional service such as timing signals, control circuits, and data distribution facilities.

The contemplated operational system must adequately render the above services and also provide the requisite system flexibility. It has been within reasonable limits by judicious sharing of channels. The principal components are:

(1) A 600 line automatic dial telephone system for administrative purposes at the launching area.

(2) A manual patch panel capable of inter-connecting 150 additional circuits for operational use.

(3) A submarine cable system connecting the down range stations providing:

  a. One voice order wire circuit providing two way communications for maintenance and servicing purposes.

  b. Twelve standard voice channels in the down range direction.

  c. Twelve standard voice channels in the up range direction.

  d. Up to six additional voice channels in the up range direction equivalent to the twelve standard up range channels except for lower signal to noise ratio.

  e. Means to block out up to nine of the standard voice channels in either direction for transmitting wide band telemetering information.

  f. Means for injecting up to eighteen duplex telegraph channels on any one of two pair of the standard voice channels.

(4) A radio network consisting of from twenty to thirty circuits at the launching area and at each down range station to provide

communication service to all remote and field stations.

In order to provide an adequate communication network prior to installation of the submarine cable, an interim system has been devised. Of necessity, the service provided is somewhat limited in extent. Provision is made to transmit point to point traffic via voice and radio teletype and broadcast timing signals for reception at down range stations. The interim equipment will be used to continually extend the range prior to the installation of the submarine cable. The same equipment will also be used in the permanent range where radio communication is required.

INSTRUMENTATION MONITORING SYSTEM:

To maintain control of the instrumentation in a range as extensive as the Florida Missile Test Range, some means of monitoring the operational status of the equipment must be included. The monitoring system provides information desired to the officers concerned.

This system consists of the transmitting and receiving panels containing the necessary switches and indicating lamps to convey the desired information. In general, all consoles in this group are composed of a standard steel executive desk, modified as required, fitted with a "turret" containing necessary lights, switches, communication equipment, etc., as required for the position concerned. Turrets consist of a housing and a front panel divided into three sections, each section composed of a standard relay rack on which is mounted the various items of equipment (telephone key boxes, radio panels, and blank panels for future use). All consoles are fitted with tumbler type electric clock showing remaining time rather than the normal elapsed time. In general all consoles will vary in the information displayed, depending on the particular staff location. A typical down range station layout is shown in Figure 14. However, all staff positions will assimilate all of the informations and present the Test Control Officer with the overall readiness of the range.

CONCLUSION:

The evolution of the instrumentation system as described has been carried forward over the past several years until it is now a well integrated complex of electronic and electro-optical equipments designed to provide proper instrumentation coverage for the Florida Missile Test Range, as illustrated by a view of the central control room at Cape Canaveral in Figure 15 and an exterior of the Building in Figure 16. In many instances this has resulted in advancing the State of the art, or in fact the establishment of a new one; Guided Missile Range Instrumentation.

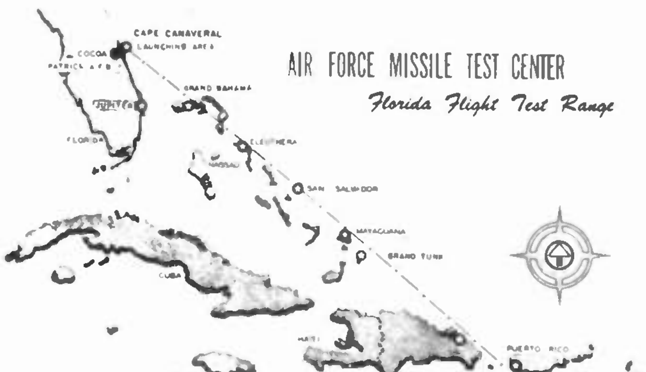


Fig. 1

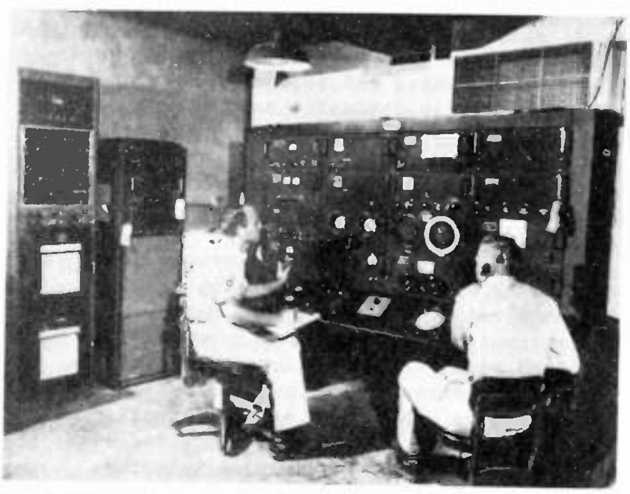


Fig. 2

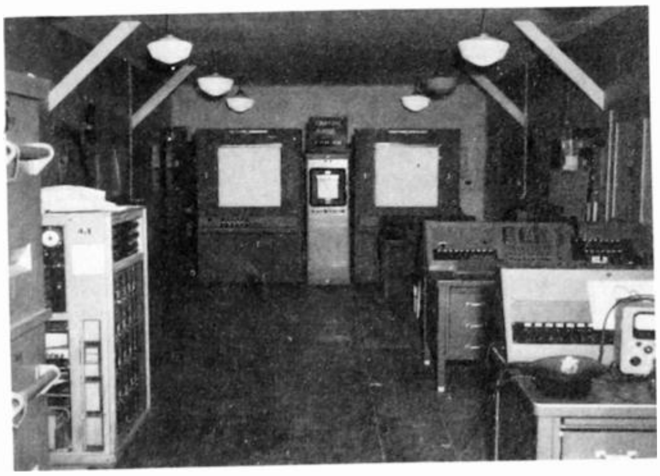


Fig. 3

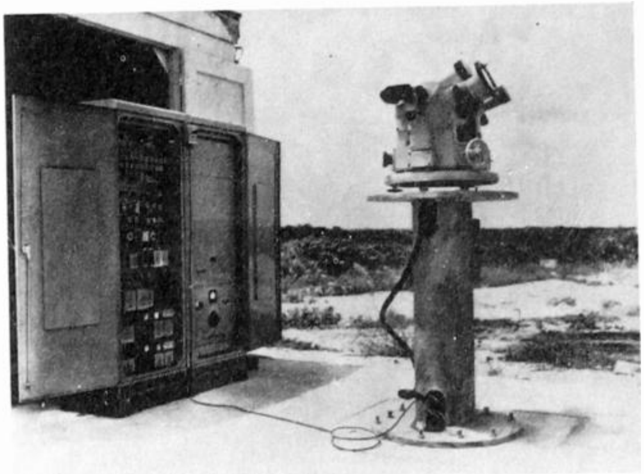


Fig. 5

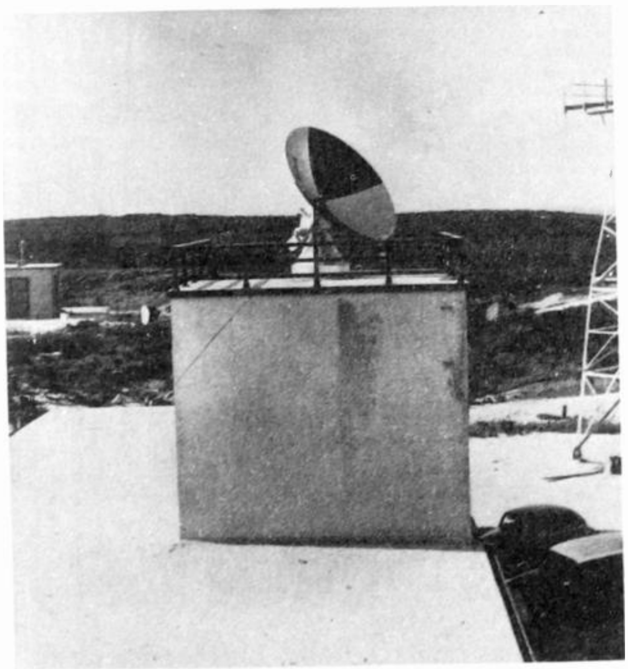


Fig. 4

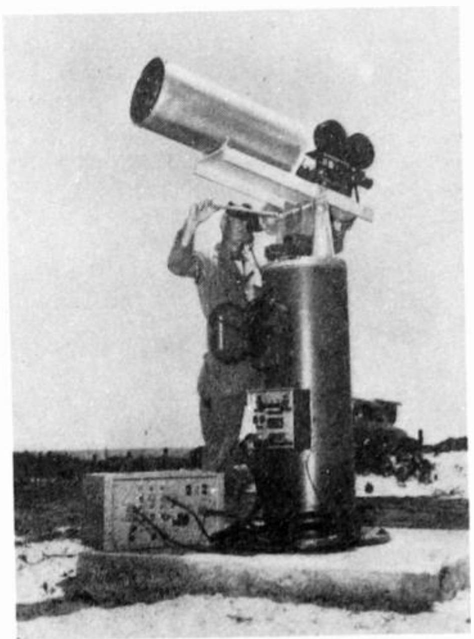


Fig. 6

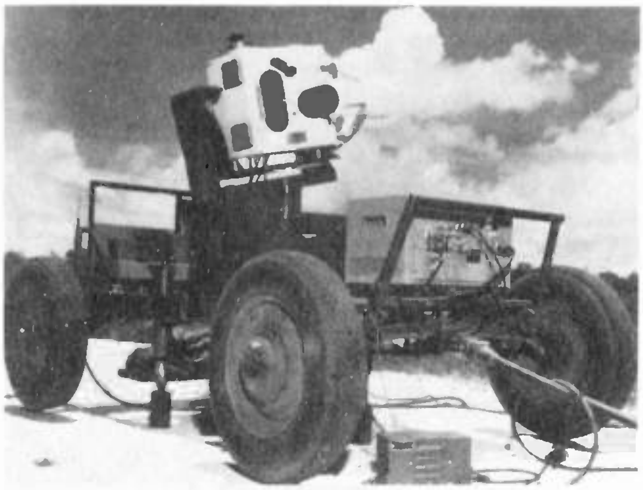


Fig. 7

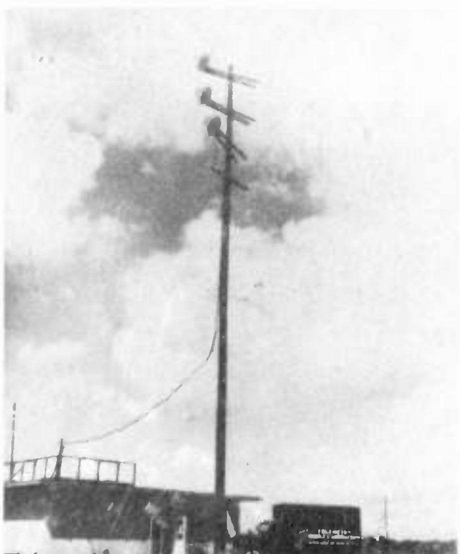


Fig. 8

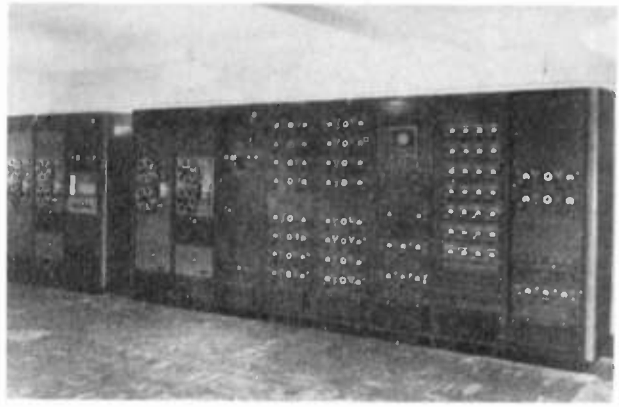


Fig. 9



Fig. 10

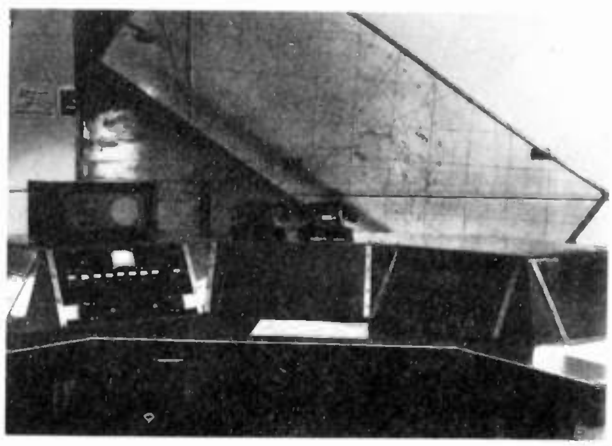


Fig. 11

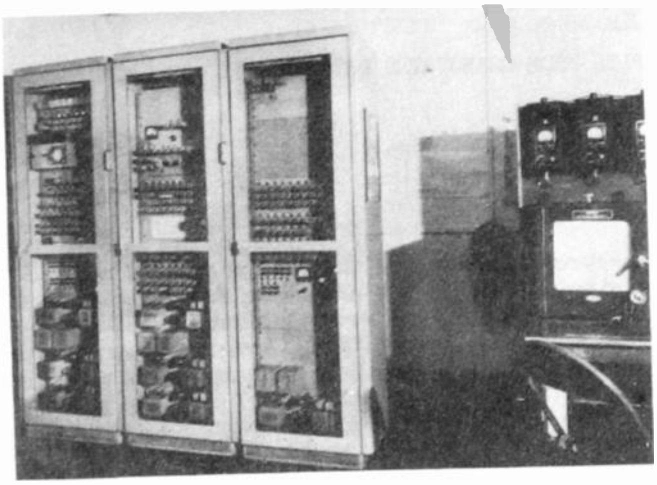


Fig. 12

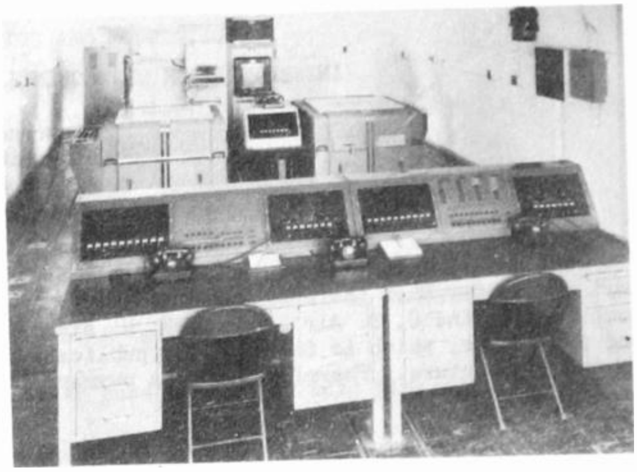


Fig. 14



Fig. 13



Fig. 15

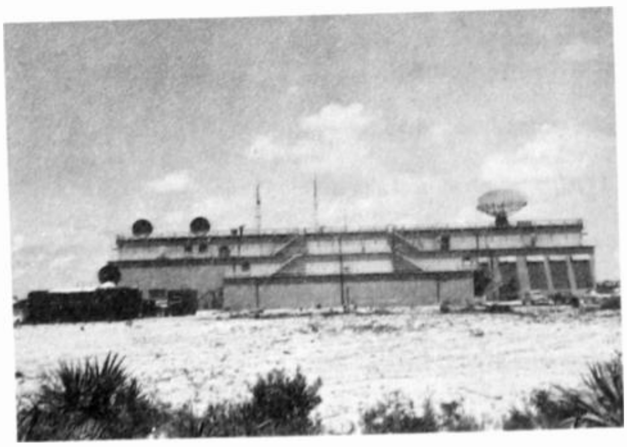


Fig. 16

INTERPRETATION OF SEQUENTIAL SAMPLES FROM COMMUTATED DATA

Lawrence L. Rauch
University of Michigan
Ann Arbor, Michigan

The material in this paper will appear as part of Chapter 9 of the book Radio Telemetry, prepared for the U. S. Air Force by M. H. Nichols and the author, which is intended for publication in the near future. Therefore, only a summary is presented here.

Summary

Most multiplexing methods used in radio telemetry are linear and come under one of the two categories of frequency division or time division. Frequency division methods make use of time-invariant filters; and if the frequency spectrum bandwidth of the data signal is larger than that for which a system is designed, the resulting error is usually one of omission where the higher frequency components of the data signal are simply lost, with some phase error introduced in the remaining components.

On the other hand, time-division methods make use of time-variant filters; and if the frequency-spectrum bandwidth of the data signal is not limited either by its nature or artificially by a low-pass filter, the resulting data samples provide ambiguous information. When an interpolation method is applied to the data samples, the resulting error is, in general, one of commission as well as omission. Frequency components are introduced which were never present in the data signal. This situation exists in many telemetering applications where it is not practical to limit the bandwidth of the data signal and where the sampling rate is limited.

The nature of the errors resulting from the ambiguity are illustrated and considered in relation to interpolation methods. Analog (non-digital) techniques which can be applied at the telemetering decommutator or during data reduction to accomplish exact or approximate interpolation are illustrated. The delay and rms error between the data signal and the interpolation signal are calculated for several representative cases in the absence of ambiguity.

COMPARISON OF REQUIRED RADIO FREQUENCY POWER IN DIFFERENT METHODS OF MULTIPLEXING AND MODULATION

By

M. H. Nichols

University of Michigan
Ann Arbor, Michigan

1. Introduction

The purpose of this paper is to present expressions for the ultimate performance of various methods of modulation and multiplexing as related to radio telemetry. Although certain of these methods may appear to be impractical with current technology, a knowledge of their ultimate performance will serve as a guide for future developments made possible by technological advances. Throughout, certain idealized assumptions will be made. However, the results can be modified to take into account departure from these assumptions in any particular case. Tables will be presented which will make possible the comparison of the various methods under threshold conditions.

2. Nomenclature

AM =amplitude modulation.

a_{oi} =unmodulated rms amplitude of ith sub-carrier (measured at the video output of the radio receiver).

α =a positive number used in connection with PAM whose reciprocal, $1/\alpha$, gives the fraction of the permissible time that an individual channel is switched on; $(1-1/\alpha)$ used to denote guard space in PDM and PPM.

β =measure of required bandwidth for FM or PM carrier in units of f_D.

β_i =measure of required bandwidth (in video) for ith FM subcarrier in units of f_{di}.

D =deviation ratio of a frequency-modulated radio link.

D_i =deviation ratio of ith FM sub-carrier.

E =information efficiency of modulation multiplex method.

F =number of samples per second per channel of a time division multiplex.

f =information frequency in an individual channel of a multiplex.

F_c =video pass band of a radio link.

f_D =maximum frequency deviation of a frequency-modulated (or phase-modulated) radio carrier.

f_{dh} =maximum frequency deviation of the highest frequency sub-carrier.

f_{di} =maximum frequency deviation of ith frequency modulated sub-carrier.

f_i =unmodulated frequency of ith sub-carrier.

f_m =maximum information frequency in a channel of a multiplex.

f_{mi} =maximum information frequency transmitted in ith channel.

k_2 =rms fluctuation noise per unit bandwidth in the video output of the comparison single channel AM link.

K_{1i} =a constant characteristic of the type of modulation of the ith sub-carrier.

M_{1i} =modulation index of the ith sub-carrier.

M_{2i} =modulation index of the radio link due to the ith sub-carrier.

M =number of binary digits in a PCM system.

PCM =pulse code modulation.

PAM =pulse amplitude modulation.

PM =phase modulation.

PPM =pulse position modulation.

PDM =pulse duration modulation.

Φ_D =maximum phase deviation of a phase-modulated radio carrier.

r =number of side band pairs in the video pass band of a PAM multiplex (i.e., video bandwidth in units of F).

R_1 =signal-to-noise ratio after demodulation.

R_2 =signal-to-noise ratio in carrier channel.

R_{oi} =wide-band gain referred to the ith channel of a multiplex.

S =rms amplitude of the sinusoidal video output of the comparison single channel AM link under condition of full modulation.

S_v =rms amplitude of carrier required for video improvement threshold in PDM, PPM and PCM and with FM sub-carriers (measured at video output of the fully modulated comparison AM link).

S_t =improvement threshold of a frequency- or phase-modulated radio carrier. (Measured at the video output of the fully modulated comparison AM link).

S_{ti} =rms amplitude of carrier required for threshold of ith FM sub-carrier (measured at video output of the fully modulated comparison AM link).

$(\frac{S}{N})_{it}$ = minimum acceptable signal-to-fluctuation-noise ratio in the output of the ith channel.

W_1 =bandwidth after demodulation.

W_2 =bandwidth of carrier channel.

3. Wide Band Gain Expressions

Wide-band gain[1] expressions for individual channel output of various types of modulation and multiplexing are given in Table 3.1. The entries in this table are taken from Nichols and Rauch[(1)] with several additions and corrections. In the frequency-division systems the wide band gain R_{oi} of the ith output channel is given by

$$R_{oi} = \frac{K_{1i} M_{1i} M_{2i}}{\sqrt{2}} \qquad (3.1)$$

The three quantities K_{1i}, M_{1i} and M_{2i} are given in Table 3.1. The $\sqrt{2}$ appears in the denominator because both side bands of each sub-carrier appear in the video of the radio receiver. Also, carrier thresholds are given. In the frequency-division systems, there are two threshold conditions: one is the rms carrier signal strength, S_t, required for carrier threshold, and the other is the rms carrier signal strength, S_{ti}, required for sub-carrier threshold assuming the carrier is above threshold. In the time-division (pulse) systems there are also two thresholds: one is the rms signal strength, S_t, required for carrier threshold and the other is the rms carrier strength, S_v,

TABLE 3.1

WIDE-BAND GAIN AND THRESHOLD EXPRESSIONS FOR VARIOUS TYPES OF MODULATION AND MULTIPLEXING

Frequency Division

	First Modulation Constant	First Modulation Index	Second Modulation Index	Carrier rms Amplitude for Sub-carrier Threshold	Carrier Improvement Threshold
Type	K_{1i}	M_{1i}	M_{2i}	$\frac{S_{ti}}{k_2}$	$\frac{S_t}{k_2}$
AM-AM	1	$\frac{a_{oi}}{S}$	1		
FM-AM	$\sqrt{3}$	$\frac{a_{oi} f_{di}}{S f_{mi}}$	1	$2.8 \frac{S}{a_{oi}} (\beta_i f_{di})^{1/2}$	
AM-FM	1	$\frac{a_{oi}}{S}$	$\frac{f_D}{f_i}$		$2(\beta f_D)^{1/2}$
FM-FM	$\sqrt{3}$	$\frac{a_{oi} f_{di}}{S f_{mi}}$	$\frac{f_D}{f_i}$	$2.8 \frac{S f_i}{a_{oi} f_D} (\beta_i f_{di})^{1/2}$	$2(\beta f_D)^{1/2}$
AM-PM	1	$\frac{a_{oi}}{S}$	Φ_D		$2(\beta f_D)^{1/2}$
FM-PM	$\sqrt{3}$	$\frac{a_{oi} f_{di}}{S f_{mi}}$	Φ_D	$\frac{2.8}{\Phi_D} \frac{S}{a_{oi}} (\beta_i f_{di})^{1/2}$	$2(\beta f_D)^{1/2}$

[1]The wide-band gain is defined as the ratio of the rms full modulation output signal-to-fluctuation-noise ratio of an individual channel of the multiplex to the rms full modulation output signal-to-fluctuation-noise ratio of a single channel AM comparison link. The received carrier power (not including side band power) of the comparison link is the same as the average received power (including modulation side bands) of the multiplex receiver. Also the rms fluctuation noise per unit bandwidth in the RF (or IF) of each receiver is the same. The output signal in each case is sinusoidal.

TABLE 3.1
(Continued)

Time Division

Type	K_{oi}	Carrier Threshold for S_t/k_2	Carrier Threshold for $S_t = S_v$
PAM-AM	$\frac{1}{n^{1/2}}$ ∞		
PAM-PM	$\frac{f_D}{(\alpha n)^{1/2}\, rF}$ ‡	$2(\beta f_D)^{1/2}$	
PAM-FM	$\frac{\pi f_D}{\alpha n F r^{1/2}}$	$2(\beta f_D)^{1/2}$	
PDM-AM	$\frac{1}{n}\left(\frac{F_c}{\alpha F}\right)^{1/2}$ Δ	$4\left(\frac{F_c}{\alpha}\right)^{1/2}$	
PDM-FM	$\frac{\sqrt{6}\, f_D}{\alpha n (F_c F)^{1/2}}$ Δ	$2(\beta f_D)^{1/2}$ *	$4\, F_c^{1/2}$
PDM-PM	$\frac{\sqrt{2}\, f_D}{\alpha n (F_c F)^{1/2}}$ Δ	$2(\beta F_c\, \Phi_D)$ *	$4.3\, F_c^{1/2}$
PPM-AM	$\frac{5F_c}{4n^{1/2}\alpha F}$ Δ	$4(nF)^{1/2}$	
PCM-AM	σ	$2(2nMF)^{1/2}$	
PCM-FM	σ	$2(\beta f_D)^{1/2}$ *	$4\, F_c^{1/2}$
PCM-PM	σ	$2(\beta F_c \Phi_D)$ *	$4.3\, F_c^{1/2}$

* Assume that the carrier threshold S_t is greater than the video threshold S_v.

∞ This is an approximate expression which holds to better than 3% if $\alpha n \geqslant 20$ and $r/\alpha n \geqslant 5/4$. See reference 1.

‡ This is an approximate expression which holds to better than 10% if $\alpha n \geqslant 20$ and $r/\alpha n \geqslant 1$. See reference 1.

Δ Assumes that there is an essentially noise free time reference for each channel.

σ The rms signal-to-rms-coding-noise ratio is given by $1.22 \cdot 2^M$. In these systems, the output channel signal-to-noise ratio is independent of fluctuation noise as long as the received signal is above threshold.

required for video slicing at half-pulse height in the pulse time systems, assuming the carrier is above threshold. The rms carrier signal strength and the rms fluctuation noise per unit bandwidth in the carrier channel (RF or IF) are expressed in terms of the output of the comparison link where S is the rms full modulation sinusoidal output amplitude and k_2 is the rms noise per unit bandwidth in the output. Thus $k_2 = \sqrt{2}\, k_1$ where k_1 is the rms noise per unit bandwidth in the carrier channel of both links. All of the symbols are defined in the Nomenclature list in Section 2.

The output signal-to-noise ratio of an individual channel can be obtained by multiplying the wide-band gain by the ratio $S/k_2 \sqrt{f_m}$ where f_m is the output bandwidth, i.e., f_m is the maximum information frequency.

4. Carrier Power Required for Minimum Acceptable Individual Channel Output Signal-to-Fluctuation-Noise Ratio

In many telemetering applications it is possible to decide on a minimum acceptable individual channel output signal-to-noise ratio $(S/N)_{it}$. It is then possible to adjust thresholds by proper choice of bandwidth so that they occur at a carrier strength which produces the required $(S/N)_{it}$. In the frequency division systems it is possible to adjust the carrier frequency deviation (in FM or PM) and the sub-carrier amplitudes and frequency deviations (in FM sub-carriers) so that the minimum acceptable $(S/N)_{it}$ for all channels occurs at the carrier theshold signal strength. In the time-division systems it is possible to adjust the carrier frequency deviation and the video bandwidth so that carrier and video thresholds occur for the same carrier signal strength and at which $(S/N)_{it}$ occurs. Table 4.1 gives the minimum rms carrier voltage to rms noise per unit bandwidth ratio (in terms of S/k_2 in the comparison AM link), required to produce the minimum acceptable individual channel output $(S/N)_{it}$.

5. Comparison of Systems

In order to compare the minimum S/k_2 required, consider a numerical case in which the minimum acceptable $(S/N)_{it}$ is 100. For the frequency-division systems, the FM sub-carrier frequencies will be based on Table I of Wynn and Ackerman[2]. Ten channels will be used, namely channels 3 through 12 in this table. These frequencies together with the information bandwidth in each channel are shown in Table 5.1. For AM sub-carriers, the sub-carrier frequency f_i in Table 5.1 will be scaled down by a factor of 5. The tapers indicated in the footnote to Table 4.1 are used and a 40% sub-carrier amplitude improvement is used as determined from Figure 4 of Nichols and Rauch[1]1.

TABLE 4.1

BANDWIDTH AND IMPROVEMENT THRESHOLDS FOR MINIMUM ACCEPTABLE OUTPUT SIGNAL-TO-NOISE RATIO $(S/N)_{it}$[1]

Type	Formulae
AM-AM	$\left(\frac{S}{k_2}\right)_{min} = \frac{\sqrt{2}\,S}{a_{oi}} f_{mi}^{1/2} \left(\frac{S}{N}\right)_{it}$
AM-FM	$f_D = \left[0.7 \frac{S f_i}{a_{oi}} \left(\frac{S}{N}\right)_{it}\right]^{2/3} \left(\frac{f_{mi}}{\beta}\right)^{1/3}$
	$\frac{S_t}{k_2} = 2(\beta f_D)^{1/2}$
AM-PM	$f_D = \Phi_D f_h = \left[0.7 \frac{S f_h}{a_{oi}} \left(\frac{S}{N}\right)_{it}\right]^{2/3} \left(\frac{f_{mi}}{\beta}\right)^{1/3}$
	$\frac{S_t}{k_2} = 2(\beta f_D)^{1/2}$
FM-AM	$D_i = \frac{f_{di}}{f_{mi}} = \left[\frac{0.28}{\beta_i^{1/2}} \left(\frac{S}{N}\right)_{it}\right]^{2/3}$
	$\frac{S_v}{k_2} = \frac{2.8S}{a_{oi}} (\beta_i f_{di})^{1/2}$
FM-FM	$D_i = \left[\frac{0.28}{\beta_i^{1/2}} \left(\frac{S}{N}\right)_{it}\right]^{2/3}$
	$f_D = \left(\frac{S f_i}{a_{oi}}\right)^{2/3} \left(\frac{2\beta_i f_{di}}{\beta}\right)^{1/3}$
	$\frac{S_t}{k_2} = \frac{S_v}{k_2} = 2(\beta f_D)^{1/2}$
FM-PM	$D_i = \left[\frac{0.28}{\beta_i^{1/2}} \left(\frac{S}{N}\right)_{it}\right]^{2/3}$
	$f_D = \Phi_D f_h = \left(\frac{S f_h}{a_{oi}}\right)^{2/3} \left(\frac{2\beta_i f_{di}}{\beta}\right)^{1/3}$
	$\frac{S_t}{k_2} = \frac{S_v}{k_2} = 2(\beta f_D)^{1/2}$
PAM-AM	$\left(\frac{S}{k_2}\right)_{min} = (n f_m)^{1/2} \left(\frac{S}{N}\right)_{it}$
PAM-FM	$f_D = \left[\frac{\alpha n F}{2\pi} \left(\frac{r f_m}{\beta}\right)^{1/2} \left(\frac{S}{N}\right)_{it}\right]^{2/3}$
	$\frac{S_t}{k_2} = 2(\beta f_D)^{1/2}$
PAM-PM	$f_D = \left[\frac{rF}{2} \left(\frac{\alpha n f_m}{\beta}\right)^{1/2} \left(\frac{S}{N}\right)_{it}\right]^{2/3}$
PDM-AM	$F_c = 0.25n\,(\alpha F f_m)^{1/2} \left(\frac{S}{N}\right)_{it}$
	$\frac{S}{k_2} = 4F_c^{1/2}$
PDM-FM	$F_c = 0.25\alpha n\,(F f_m)^{1/2} \left(\frac{S}{N}\right)_{it}$
	$f_D = 0.4F_c,\ \beta = 10$
	$\frac{S_t}{k_2} = \frac{S_v}{k_2} = 2(\beta f_D)^{1/2}$
PDM-PM	$F_c = 0.25\alpha n\,(F f_m)^{1/2} \left(\frac{S}{N}\right)_{it}$
	$\Phi_D = 0.6,\ f_D = 0.6F_c,\ \beta = 8$
	$\frac{S_t}{k_2} = \frac{S_v}{k_2} = 2(\beta f_D)^{1/2}$
PPM-AM	$F_c = 0.2\alpha n\,(F f_m)^{1/2} \left(\frac{S}{N}\right)_{it}$
	$\frac{S_t}{k_2} = 4(nF)^{1/2}$
PCM-AM	$F_c = \frac{7nF}{2}$ for a code of 7 binary digits, i.e., $M = 7$
	$\frac{S_t}{k_2} = 7.5\,(nF)^{1/2}$
PCM-FM	$f_D = 1.4\,nF$ for $M = 7$
	$D = 0.4,\ \beta = 10$
	$\frac{S_t}{k_2} = \frac{S_v}{k_2} = 7.4\,(nF)^{1/2}$
PCM-PM	$f_D = 2.1nF$ for $M = 7$
	$\Phi_D = 0.6,\ \beta = 10$
	$\frac{S_t}{k_2} = \frac{S_v}{k_2} = 8\,(nF)^{1/2}$

[1] In the frequency-division systems the formulae are in terms of the sub-carrier which requires the highest carrier strength for threshold. If $(S/N)_{it}$ is the same for all channels and if f_{mi} is proportional to f_i for all sub-carriers and if f_{di} is proportional to f_i for FM sub-carriers, then for threshold to occur simultaneously for

The total information bandwidth of the ten channels is 1085 cps. For the time-division systems, ten channels with a bandwidth of 100 cps each will be assumed, i.e., $f_m = 100$. Also the values $F = 250$, $\alpha = 2$ for PAM and $\alpha = 1.2$ for PDM and PPM will be used. Table 5.2 gives the minimum S/k_2 in each case, the carrier power relative to PPM-AM and the required RF bandwidth. In the case of PCM a seven digit code is used for which the rms full-modulation output signal-to-rms-coding-noise ratio is 160.

TABLE 5.1

SUB-CARRIER CENTER FREQUENCIES
AND OUTPUT BANDWIDTHS

f_i	f_{mi}
1300	20
1700	25
2300	35
3000	45
3900	60
5400	80
7350	110
10500	160
14500	220
22000	330

6. Information Efficiency

The information efficiency E is defined by

$$E = \frac{W_1 \log (1 + R_1^2)}{W_2 \log (1 + R_2^2)} \qquad (6.1)$$

where W_1 is the total output bandwidth (1000cps in the numerical case), W_2 is the RF bandwidth, R_1 the output signal-to-noise ratio in each channel of the multiplex (assumed to be the same for each channel) and R_2 the signal-to-noise ratio in the RF[1]. The information efficiency E is a measure of the effectiveness of the exchange of bandwidth for signal-to-noise ratio in the wide band modulation of the subcarriers and/or RF carrier. This quantity is particularly interesting for the minimum acceptable output signal-to-noise ratio. Table 6.1 gives E for the various systems at the condition of minimum acceptable output $(S/N)_{it}$ and also at threshold conditions when they apply. This table is based on Table V of Nichols and Rauch(1) (plus some additions) and on the numerical values in Table 5.2 for which $R_1 = (S/N)_{it} = 100$ for all except the PCM methods in which $(S/N)_{it} = 160$. It should be noted that for the minimum acceptable output signal-to-noise ratios stated, PCM-AM and PPM-AM have about the same information efficiency. But at higher minimum acceptable $(S/N)_{it}$, the PCM methods will eventually have higher information efficiency than the analogue types of modulation. However, in radio telemetry, the environmental errors such as vibration, etc., are usually the order of a percent or more, so that $(S/N)_{it}$ values larger than several hundred or so are usually not very important. Of course, in the analogue system, the output S/N increases linearly with increasing signal strength (when above threshold) so that the output S/N can be arbitrarily large. In PCM methods the output S/N is essentially constant independent of received signal strength when above threshold.

(Continued from previous page) all sub-carriers, the sub-carrier amplitudes must be tapered. With an FM carrier, the sub-carrier amplitudes must be proportional to $f_i^{3/2}$ and with an AM or PM carrier the taper is $f_i^{1/2}$. In the time-division systems the formulae are in terms of the channel with the largest $(S/N)_{it}$ it being understood that all channels are sampled at the same rate, etc.

[1] This is an approximation since the sub-carrier amplitudes in this illustration are tapered whereas Figure 4 of Nichols and Rauch applies to sub-carriers of same amplitude.

TABLE 5.2

MINIMUM S/k_2 FOR $(S/N)_{it} = 100$
THE CORRESPONDING RF BANDWIDTH AND POWER RELATIVE
TO PPM-AM

Type	$\frac{S}{k_2}\Big]_{min}$	Power Relative To PPM-AM	RF Bandwidth In Kilocycles
PPM-AM	200	1	76
PCM-AM	370	3.4	18
PCM-FM	370	3.4	35
PCM-PM	400	4.0	42
PAM-FM	580	8.3	85
AM-FM	600	9.0	92
FM-FM	680	11	115
AM-PM	760	14	145
PAM-PM	780	15	150
FM-PM	800	16	160
FM-AM	820	17	50
PDM-AM	830	17	88
PDM-FM	870	19	188
PDM-PM	980	23	245
PAM-AM	3150	250	18
AM-AM	9600	2300	9.5

7. Double Multiplexing

In many radio telemetering applications, a

TABLE 6.1

INFORMATION EFFICIENCIES

Method	E	RF Bandwidth in Kilocycles
PCM-AM	0.21	18
PPM-AM	0.17	76
PCM-FM	0.13	35
PCM-PM	0.11	42
PAM-AM	0.080	18
FM-AM	0.055	50
PAM-FM	0.050	85
AM-FM	0.046	92
PDM-AM	0.037	88
FM-FM	0.037	125
PAM-PM	0.029	150
AM-PM	0.028	155
FM-PM	0.027	175
PDM-FM	0.022	188
PDM-PM	0.017	244

group of relatively high frequency response channels is required plus a group of relatively low frequency response channels. For technical reasons, this has been accomplished by "subcommutation." In frequency division multiplexing, it is necessary to provide sufficient bandwidth in the sub-carrier channel which is commutated to prevent objectionable overlapping of the pulses. This requires a bandwidth of about 3.5 n' F' to 4 n' F' where n' is the number of channels in the subcommutator and F' is the number of samples per second per channel. In a time-division multiplex, the subcommutator can be synchronized with the main commutator so that one channel of the subcommutator is sampled in each cycle of the main commutator, or the subcommutator can be run without synchronization. Without synchronization, a bandwidth of 4 n' F' is required as in frequency division. However, it takes about 10 times as much bandwidth for the unsynchronized case as for the synchronized. The reason for this is as follows. Suppose that in the main commutator the low-pass output filter cuts off at the frequency f_m = F/2.5. To keep the pulses from over-lapping, f_m = 4 n' F'. Thus the total sampling rate n' F' of the subcommutator is related to the sampling rate F per channel of the main commutator by

$$n' F' \text{ (unsynchronized)} = F/10 \quad (7.1)$$

But if the subcommutator is synchronized, then

$$n' F' \text{ (synchronized)} = F \quad (7.2)$$

Thus if F is held constant, the synchronized subcommutator can have n' F' ten times larger than in the unsynchronized, i.e., ten times more information bandwidth. It is easy to see that if the synchronized subcommutator uses the same number of samples per cycle of information as the main commutator, then the output S/N will be the same in the subcommutator as in the main commutator.

As a numerical example, let n' F' = 100. Of the frequency-division methods only FM-FM will be considered. With a subcommutator this becomes PAM-FM-FM. If n' F' = 100, then a sub-carrier bandwidth of around 350 to 400 cps is required to keep crosstalk down. Suppose that the 22 kc channel from Table 5.1 is used. Then by combining the expressions for PAM and FM-FM wide-band gains, the wide-band gain for PAM-FM-FM can be determined. Then, based on Table 5.2 the output S/N of the subcommutated channels can be computed under threshold conditions with S/N of the sub-carrier outputs equal to 100. The result is given in Table 7.1. Also in Table 7.1, the output S/N of the subcommutated channels is also given for time-division systems with the synchronized and unsynchronized subcommutators. These results are also based on Table 5.2. All unsynchronized subcommutation is taken as PAM since this requires the least bandwidth. All synchronized subcommutation has the same type of modulation as the main multiplex. For the PCM systems with unsynchronized commutator, it has been assumed that the coding noise in the output of the main commutator is "white". For the unsynchronized case, four channels of the main commutator are required, i.e., 400 cps of bandwidth are required. For the synchronized case, n' F' = 250 is available.

TABLE 7.1

COMPARISON OF OUTPUT SIGNAL-TO-NOISE RATIO OF DOUBLE MULTIPLEXING (SUBCOMMUTATION) AT AT THRESHOLD OF MAIN MULTIPLEX[1]

System	Unsynchronized	Synchronized
PAM-FM-FM	440	
PAM-PAM-FM	220	
PAM-PDM-FM	220	
PAM-PPM-AM	220	
PAM-PAM-FM		160
PDM-PDM-FM		160
PPM-PPM-AM		160
PCM-PCM-AM		260
PAM-PCM-AM	360	

[1] See Table 5.2

So if 20 subcommutated channels are used, then F' = 12.5. The entries for the synchronized case are based on these figures.

It should be pointed out again that in the unsynchronized case, ten times as much bandwidth is required for crosstalk reasons, etc.; consequently the output S/N are larger than for the synchronized case.

8. Assumptions

Throughout, the RF, IF, and video passbands have been assumed as having ideal sharp-cutoff characteristics and the RF and IF passbands have been assumed just wide enough to accommodate the modulation sidebands. Of course, in any practical case, effective noise bandwidths must be substituted because of the impossibility of realizing ideal sharp cutoff and because leeway has to be provided for drifts in frequency of the transmitter and receiver oscillators, etc. The same may be said for the sub-carrier channels in a frequency-division multiplex. In the time-division systems, it has been assumed that a noise-free time reference is available for operating the receiver commutator and timing circuits, etc. Allowance can be made for noise on the time reference in any particular case.

A lengthier discussion of the material presented in this paper will be found in the book Radio Telemetry by M. H. Nichols and L. L. Rauch. This book is being written for the U. S. Air Force and will be published in the near future.

REFERENCES

(1) M. H. Nichols and L. L. Rauch, Rev. Sci. Instrum. 22 1, (1951).

(2) J. B. Wynn and S. L. Ackerman, Electronics vol. 25, 106 (May 1952).

FLIGHT TESTING OF AN AIRBORNE DIGITAL COMPUTER*

E. M. Grabbe and D. W. Burbeck

The Ramo-Wooldridge Corporation
Los Angeles, California

and

S. B. Neister, Armament Laboratory

Wright Air Development Center
Dayton, Ohio

Summary

Flight tests of the first airborne digital computer have been highly successful. A general purpose digital computer was flight tested in a C-47 aircraft as part of the Digitac Airborne Control System. The computer automatically controlled the aircraft in a number of flights through a series of four way points, and automatic control was smoother and more accurate than manual control. Details of the flight tests are reported and the demonstrated advantages of digital computers for airborne control applications are discussed. These flight tests have clearly demonstrated that airborne digital computers are practical and add great versatility and flexibility to an automatic control system.

Introduction

Electronic digital computers have certain unique advantages, such as accuracy, flexibility, and versatility, which make them ideally suited for use in many military airborne control systems. However, for airborne use a digital computer must meet the additional stringent requirements imposed by military aircraft of size, weight and facilities for cooling, beyond the requirements normally encountered in digital computer design. The past few years have seen many improvements in computer logics, circuitry, and packaging, so that compact general purpose airborne digital computers have now become a reality.[1] The Digitac System was the first in which an electronic digital computer was employed as an integral part of an airborne automatic control system. This paper contains a description of the evaluation tests of the Digitac System and reports certain unclassified results of these tests.

Digitac, standing for Digital Tactical Automatic Control, is a military system developed for completely automatic aircraft navigation and weapons control at line-of-sight ranges. This report of the Digitac flight tests will cover (a) a description of the system and the equipment developed for it, (b) the flight test objectives and details of the flight tests, and (c) a discussion of advantages of digital computation for future airborne control application. The weapons control part of the system has a security classification and, hence, will not be discussed. A detailed description of the computer used in these flight tests is being presented in another paper at this convention.[2]

*The work reported in this paper was done at the Hughes Aircraft Company, Culver City, California.

Description of System and Equipment

The Digitac Airborne Control System[3] was developed under contract with the Armament Laboratory, Wright Aeronautical Development Center. The objectives of the program were to develop a high-precision navigation system based on hyperbolic position determination. In a hyperbolic system two pairs of ground-based transmitter stations are required to establish the coordinate system. Usually one of the stations is common so that three stations are sufficient: a master station and two slave stations. The master station alternately sends pulses to each of two slaves which relay the pulses to the aircraft, and measurements of the differences in time of arrival of pulses from the master station and the two slave stations define two hyperbolic lines of position so that a fix may be determined by the intersection. Hyperbolic systems have an inherent error factor due to divergence of the coordinates, which increases with range from the ground stations. Hence, for precision navigation, high accuracy in both time measurements and computation are required.

Equipment developed for the Digitac System consisted of one set of three ground stations and two models of the airborne system. The airborne equipment, shown in Figure 1 in block diagram form, consists of a receiver, decoder and time measurement equipment, instrument analogue-to-digital conversion units, an input register, a digital computer, and outputs. The time measurement part of the system was designed to automatically measure time in digital form with an over-all accuracy of one part in 30,000 or about ± .02 microseconds.[4] The aircraft instrument inputs to the system consisted of altitude, air speed, and compass heading, and a counter type shaft to digital converter was used for each of three instruments inputs.[5] The outputs consisted of a steering signal, a pilot's display and certain on-off signals.

Figure 2 shows the computer model which was flight tested. The general purpose computer is a serial magnetic drum machine

with a word length of sixteen binary digits plus sign and a digit frequency of 100 kc. Figure 3 is the input-output equipment including the input register, time measurements circuits, instrument input converters and output units.

The aircraft used in these flight tests was a cargo type C-47 in which an E-6 autopilot was installed. A vertically stabilized camera was used in the aircraft to determine position so that the accuracy of flight control could be determined. This aircraft proved to be an ideal flying laboratory in addition to being a reliable craft.

Objectives

The objectives of the flight test program were to evaluate the flight performance of this navigational system including the digital computer. Considerable care was taken to lay out a test program that would provide the maximum operational data on the system. It should be pointed out that this constituted the first flight test of an airborne digital computer and the program was approached with the usual concern which accompanies the first trial of something completely new.

The following specific objectives were established for the flight test program:

1. To determine the accuracy of the position determination of the system by making measurements on the ground;
2. To determine the accuracy of position determination in flight using the vertically stabilized camera in the aircraft;
3. To carry-out navigation through a series of way points, first by (a) pilot operation from PDI signals, and second (b) by automatic control of the aircraft through a digital autopilot coupler;
4. To determine the system accuracy of automatic navigation and dead reckoning.

Flight Tests

The three ground stations required for the system tests were located on hilltops in the West Los Angeles Area and formed two baselines five to six miles in length. This setup provided a sizeable test area in the neighborhood of Metropolitan Los Angeles where there are excellent check points for both ground and air tests, since many street corner locations are known to first order survey accuracy.

For the first objective of determining the accuracy of fix on the ground the receiver and time measurements equipment without the computer were installed in a van for mobile tests. Digital time measurements were recorded at a number of accurately known points, and position was computed to compare with geodetic coordinates. It should be noted that while most of the data concerning these system tests have been declassified the operational data involving accuracies of the system still have a security classification.

The flight test program for completing the other three objectives was carried out over a period of fourteen months; fifty-nine flights were made for a total of ninety-two hours of flight time. This amounted to about one and one-half hours of flying time per week during the period of the test program. Of these flights, approximately three-quarters were made using the airborne digital computer.

Part of the test equipment developed for the flight test program was a digital data recorder. Fifteen digital quantities consisting of input data and results of computation could be extracted from the computer on call, displayed on a cathode ray tube and photographed. Data could be recorded at the rate of about two sets per second, automatically or at longer intervals under manual control. At 150 m.p.h., a reasonable cruising speed for the C-47, the aircraft travels about 220 feet per second; hence, data could be recorded about every 100 feet. This test equipment proved very valuable in trouble-shooting and checking out of components as well as in evaluating the performance of the complete system.

In carrying out the second objective of determining accuracy of position in the air, flights were made over well-defined locations close to first order survey points in the area; for example, the Santa Monica water tower is near a first order survey point and has the advantage that the diameter of the water tank gives an additional yardstick for measurements in the photographs. The time measurements were recorded when the aircraft was directly over the selected point and simultaneously a photograph was taken with the vertically stabilized camera. The accuracy of position determination could then be obtained by comparing these two records. After the accuracy of the time measurements had been established from the vertical photographs, the position of the aircraft in later flights was determined mainly from the recorded time measurements data.

For completing the last two objectives of the system tests a complete programing of the system equations for navigation was required. A flow chart of the computer program is shown in Figure 4. From the time measurements and instruments data, position is determined in X-Y-Z coordinates, with the X and Y axes corresponding to N-S and E-W directions. A fast-converging iterative method involving predicted position was used for transforming hyperbolic to rectangular coordinate position. The wind is determined from the computed ground velocity and the measured true air speed. A steering signal is then computed which provides a heading correction for the aircraft to direct it to a selected destination. In the absence of time measurement signals, dead reckoning was carried out using a ground velocity computed from the previously computed wind and the measured air speed. In the final stages of the flight test program the computer determined smoothed position, smoothed wind, ground velocity and steering signal. The program, consisting of approximately 350 steps, required about .5 of a second second. The computation time for the dead

reckoning mode of operation was .3 of a second, and since the iteration time was held fixed at .5 of a second there remained 0.2 of a second dead time. By periodically forcing the computer into the dead reckoning branch of the program, computations for the weapons control part of the system were solved during this additional 0.2 of a second. Hence, many auxiliary computations can be carried out without increasing the length of computation time or decreasing accuracy of control.

Figure 5 shows the manner of connecting the coupler to the autopilot system. The upper loop shown in this figure without the differential is the normal loop for autopilot control, while the lower loop modifies the autopilot heading through a differential for automatic digital computer control. This approach provides control without interfering with the normal operation of the autopilot loop. Details of the autopilot coupler used in the system are being reported in another paper at this convention.[6]

For checking out the program in the laboratory before flight testing, an automatic plotting board was connected with the computer, making it possible to "fly" in the laboratory. This was done by entering a position, wind, air speed and heading into the computer together with the coordinates of a sequence of way points, and the dead reckoned course of the computer through these way points was plotted on the board. The output steering signal was used to modify the heading input through the autopilot coupler as if the aircraft had zero response time. This technique was valuable in determining errors in the program and also provided an excellent laboratory demonstration.

After laboratory checkout had been completed the third objective of navigating the aircraft through a series of four way points was carried out. Four accurately known locations in the West Los Angeles Area are shown in Figure 6 together with the baselines. The point nearest the baseline is a roofhouse at Hughes Aircraft Company. The other points are the Santa Monica water tower, a street corner in the San Fernando Valley and a radio tower in the Santa Monica Mountains. This group of points covers the geometrical cases of high, medium, and low divergence of hyperbolic coordinates and one point behind the baseline.

Initial flights through this series of way points were made by having the pilot control the aircraft from output signals displayed on a Pilot's Direction Indicator. These flights with manual control were highly successful and many improvements in programing weremade on the basis of the operational results. They also served to provide a basis for comparison with completely automatic control. The steering signal from the computer is such as to guide the aircraft in a straight-line path to a selected destination from any point in the service area. In this manner the aircraft is guided to the first point of the sequence from any starting point. On reaching a pre-selected range from this point the computer selects the next point in the sequence, etc. On reaching the fourth point the aircraft is then again directed to the first point.

This ability to program a computer through a sequence of way points is an extremely important characteristic which should prove very valuable in operational automatic flights. It is estimated that twenty to thirty way points could easily be programed in the computer memory in this fashion.

One of the big steps in the flight test program was the first completely automatic flight around this same array of way points. The autopilot coupler was connected and the flight through the first way point was made under pilot's control. The system was then switched to automatic control and it successfully guided the aircraft through the sequence of way points. In comparison with manual control the automatic control was smoother and more accurate. The computer iteration time of .5 seconds did not lead to any difficulties in stability; in fact stable control was obtained in later flights with longer iteration times.

For the final objective of determining the accuracy of automatic navigation and dead reckoning, the digital data recorder described earlier in this paper was used extensively. For automatic navigation, data were recorded at .5-second intervals and the control accuracy was determined by comparing computed position with position derived from raw time measurement data for points along the flight course. Vertical photographs were also taken in some tests. Dead reckoning accuracy was also very easily checked by forcing the computer into this mode of operation and simultaneously recording dead reckoned position and raw time measurement data for comparison. As indicated previously, quantitative results of these accuracy tests are classified.

It should be noted that one of the way points in Figure 4 lies behind the baseline in a region where some position ambiguity might be encountered; however, the computation technique, using predicted position, permitted accurate flight through this area. The point in the Santa Monica Mountains lies in a region of the hyperbolic field where the coordinate divergence is high; however, the programing for the problem also gave smooth control in this area.

Many changes were made in the programing of the computer during the course of the flight test program in order to improve the system operation. Constants in the program may be readily changed as well as program steps. For example, various constants were tried in flight to determine optimum velocity smoothing, and the final choice was to smooth by taking one-half the position computed from time measurements and one-half the predicted position. Dead reckoned position was computed from the smoothed wind and air speed vectors, and during dead reckoning no changes were made in the wind. In smoothing the wind 1/32nd of the computed wind was combined with 31/32nds of the previous smoothed wind to obtain a new value. This provided a heavy damping on the

wind computation and 20 to 30 seconds are required to accumulate a stable wind. It is not required that the aircraft fly a straight, level course during this settling time, but only that coordinate information be received during this period. The constants chosen for smoothing the aircraft position and wind are for this particular aircraft-autopilot combination and in general will be different for other systems.

All of these programing changes indicate clearly the value of being able to make quick changes in the problem without changes in equipment. This is especially valuable in flight tests for system evaluation in which delays may ground an aircraft with a field test crew standing by.

Conclusion

These flight tests have shown that the use of a digital computer in an airborne control system is practical and that it adds tremendous versatility and flexibility to an automatic control system that would be difficult to obtain by other techniques. The unique advantages of a digital computer for flight control as demonstrated by these tests are as follows:

1. High computational accuracy is easy to obtain with a digital computer. This is especially important since there is a tendency toward higher accuracy in new systems.
2. The problem formulation can be readily changed without changes in the computer equipment. This also means that one airborne digital computer may be used in a variety of different airborne control systems.
3. A programed flight through a selected series of way points is possible. Up to 20 or 30 way points may be used and a selector switch may be provided to select any destination or sequencing may be followed automatically.
4. The equipment is easy to mass produce, since essentially no precision parts are required. The use of plug-in sub-assemblies not only makes for ease of fabrication but also simplifies maintenance, repair, and spare parts problems.

In looking to the future we may predict that the day is not too far-off when one standard airborne digital computer will be used in a variety of different military and commercial systems for automatic navigation and control. Such a digital computer in an aircraft can not only handle the control functions of the system but may also act as the central computing point for a variety of auxiliary computations, such as: cruise control, air data reduction, predicted time of arrival, and other problems.

In concluding it should be stated that these flight tests have been a milestone in the history of airborne digital computers. The computer equipment used in these tests reflects the state of the design art three to four years ago and is now obsolete. However, the big step has been taken to demonstrate that airborne digital computers are practical and that they have such a versatility and flexibility that future use of airborne digital computers will be wide-spread.

The program described in this paper was made possible through the cooperative efforts of a small group of engineers at Hughes Aircraft Company and the sponsoring agency, the Armament Laboratory, Wright Air Development Center.

---

[1] E. C. Nelson, "A Digital Computer for Airborne Control Systems", 1952 IRE Western Electronic Show and Convention, Long Beach, California.

[2] E. E. Bolles, "The Digitac Airborne Digital Computer".

[3] D. W. Burbeck, E. E. Bolles, W. E. Frady and E. M. Grabbe, "The Digitac Airborne Control System", Western Computer Conference, Los Angeles, California, February 1954.

[4] D. W. Burbeck and W. E. Frady, "Precision Automatic Time Measurement Equipment", 1952 IRE National Convention, N. Y. C.

[5] A. D. Scarbrough, "An Analogue to Digital Converter", 1953 Western Electronic Show and Conference, San Francisco, California

[6] W. L. Exner and A. D. Scarbrough, "A Digital Autopilot Coupler".

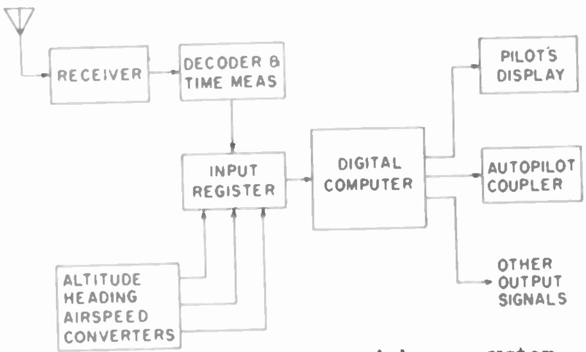


Fig. 1 - Block diagram - airborne system.

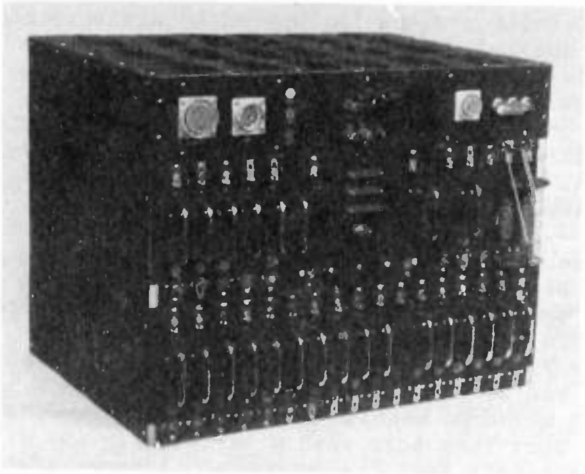


Fig. 2 - Airborne digital computer.

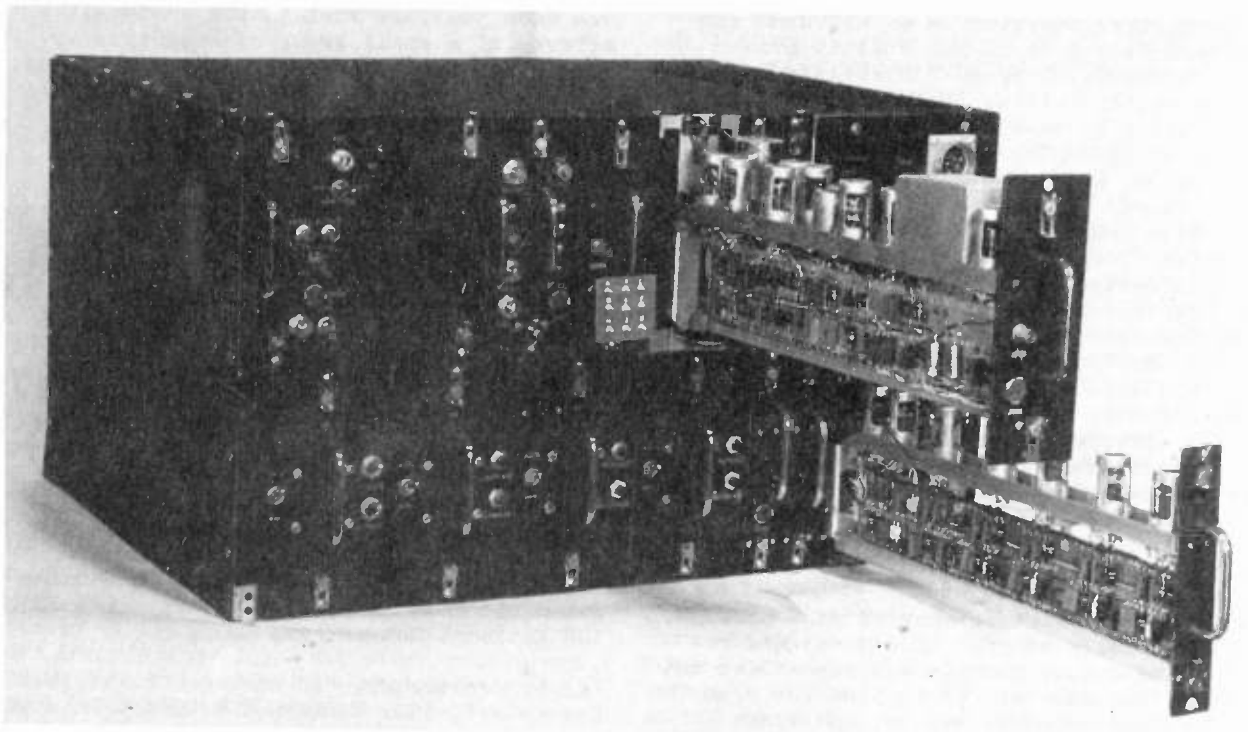


Fig. 3 - Input-output.

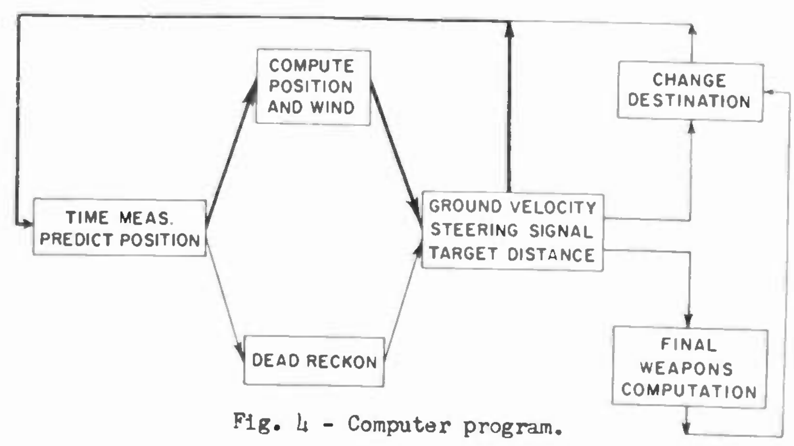


Fig. 4 - Computer program.

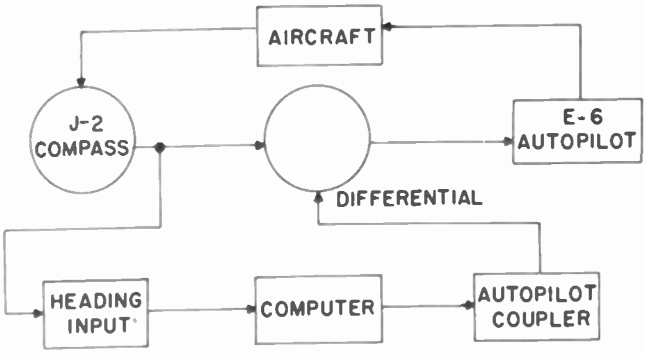


Fig. 5 - Digital computer - autopilot connection.

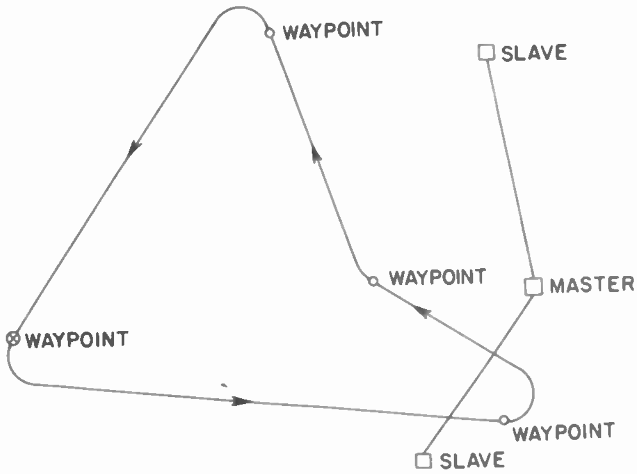


Fig. 6 - Ground stations and waypoints.

EVALUATION FOR MAGNETIC TAPE EQUIPMENT FOR TELEMETERING INSTRUMENTATION

R.E. Rawlins
Electronic Research Laboratory
Lockheed Aircraft Corporation
Burbank, California

(Abstract was not available at the time of publication.)

THE DIGITAC AIRBORNE DIGITAL COMPUTER*

E. E. Bolles
The Ramo-Wooldridge Corporation
Los Angeles, California

Summary

The first airborne digital computer to be employed in an automatic flight control system is described. This computer was developed to be both rugged and compact and yet to preserve the desirable operational characteristics of large-scale digital computers. The computer described is a serial general purpose machine having a magnetic drum memory. A floating address reference scheme is utilized to aid in minimum access time programing. The programing and coding of the system equations is outlined, including the problems of accurate position determination, data smoothing, weapons control, and automatic flight control. The operational advantages of using an airborne digital computer to solve problems of this type have been demonstrated by a series of system flight tests.

Introduction

Digitac is a completely automatic navigation and weapons control system incorporating the first airborne digital computer. The development of the digital computer for this system was undertaken in 1949, a time when the computer field was almost totally concerned with the design and construction of large-scale scientific computers. It was felt that a general purpose digital computer could be developed that would be compact enough to be airborne, thus making the accuracy and versatility of digital computation available to an aircraft control system. Such an airborne digital computer has been developed and flight tested. Because this computer was a general purpose machine, it was not only capable of accurately solving the required navigation and weapons control problems, but also demonstrated the advantages of a control element whose method of operation could be rapidly and easily changed. The ability of a digital computer to select alternate modes of operation, based on numerical data, was utilized repeatedly.

The problem of smoothly controlling an aircraft with the quantized steering information supplied by the computer is discussed in another paper being presented at this convention.[1] The results of the flight tests of this system are also being presented at this convention.[2] It suffices to say here that the flight tests were successful in proving not only that airborne digital computers can accurately solve the complex problems of aircraft control, but also that they can provide numerous advantages never before realized.

Design Features

The aircraft's position in the Digitac system is determined by measuring the difference in time of arrival of pulses transmitted by two pair of ground stations. These pulse pairs define two hyperbolas, the intersection of which is the aircraft's position. The required navigation accuracy of this system indicated the need of developing a new method of automatically measuring these time differences. The measuring system is basically a counting technique combined with a vernier interpolation scheme. The equipment based on this principle is capable of measuring time differences to an accuracy of .02 microseconds, or about one part in 30,000. To take full advantage of this accurate digital input data it was decided that digital computation was necessary. Further investigation indicated that though the accuracy consideration is important, actually the versatility of digital computers is their predominent advantage.

Since this computer was to be a part of an airborne system and thus be subjected to the rigors of an aircraft environment, many of the basic computer design considerations were predetermined. The requirement of minimum size and weight indicated that the computer be a straight binary machine with operations performed in a serial manner. An early analysis of the system accuracy led to the use of numbers consisting of 16 binary digits plus sign. The choice of a general purpose design, with the problem instructions stored in the computer memory, was made to facilitate the many changes it was felt would arise during the flight test program. This internal program storage also allows the computer to be readily adapted to handle other system problems and new field situations.

*The work reported in this paper was done at the Hughes Aircraft Company, Culver City, Cal.

The memory of the computer is a rotating magnetic drum, chosen principally because it is a non-volatile storage system. Power and equipment failures, therefore, cannot destroy the memory contents. The magnetic drum memory conforms to the need for minimum size, since the information stored per unit volume is very high. This type of memory is also well suited to use in a serial computer.

The memory has storage space for 768 two-address orders and 192 constants. Ninety-six storage locations and a six-word, fast access circulating register are available for the storage of the results of computation.

The arithmetic element of this computer is similar in logic to most present serial machines. It is composed of three static registers, an adder, and a sign control unit. Numbers are stored in the memory as absolute value plus sign. Subtraction is accomplished by adding complements; therefore, the sign of each number transferred to or from the memory is sensed and if necessary the number is complemented during the transfer operation.

The order code of the machine consists of 37 operations; 24 of these direct the sampling of inputs and ordering of outputs for navigation and weapons control. The remaining operations are the normal arithmetic operations of addition, subtraction, multiplication, division, and the associated transfer operations. To achieve minimum solution time, the arithmetic element is mechanized so that the result of any operation may be used as the first operand of another operation without the necessity of storing this result in the memory. To accomplish these series or cumulative operations, special addition, subtraction, multiplication, and divisions orders are provided.

One decision operation is provided which is used to select one of two alternate successive orders dependent upon the sign of the number in the arithmetic element. The use of this type of order, which is quite common in digital computers, provided a system capable of selecting alternate modes of operation based on prevailing conditions, rather than predicted conditions.

To achieve minimum problem solution time it was felt desirable to have a two-address, rather than single-address, code for the control element. In order to save memory space it was necessary to compress the information consisting of an operation, an operand address, and the next order address into a single word of 17 digits. Figure 1 shows the digit designation for an order word. Four digits of the word are used to designate the memory band (head) of the operand. This same band information also designates the band from which the next order will be obtained. Order and number bands are physically separate heads so the same band number can designate both a number and an order head. Four binary digits are used to designate the operation to be performed. Of the 37 operations performed by the machine, only 13 are arithmetic operations requiring reference to the memory. Since the four binary digits can designate a maximum of 16 operations, only three operations remain for input and output control. These operations, however, do not require memory reference so the information normally designating the band of the operand is not required. These three basic operations are each modified by three of the band digits giving a total of 24 input-output operations.

The last nine digits of the order word are used to complete the designation of the operand and the next successive order. As previously mentioned, the band of the operand and the next order are specified so it is only necessary to indicate the desired sector. In most magnetic drum machines a sector is specified by a number given with reference to an origin or zero location on the magnetic drum. In the Digitac computer the address reference is not fixed, but rather the address is referenced to the completion of the last order operation. The ninth through thirteenth digit of the order word are the number of sectors between where the present order is read and the location of the operand. After the order is read it is only necessary to count the indicated number of sectors and then perform the operation. The last four order digits are the number of sectors to count after completing the operation until reading the next order word.

This floating reference addressing system not only allows a compact order word, but also is a great aid in minimum time programing since the coder is aware of the access and operation time for every order in the program sequence.

Since each step of the order sequence is dependent upon the previous step, random errors may cause one of two types of sequence failures. An order may be selected from a blank address in the memory; in this case the sequence will immediately return to the beginning of the routine. If the selected address does contain an order, then the sequence will continue from that point. Both of these failures may cause discontinuities in the output; however, the filtering action on the outputs will minimize these discontinuities and the outputs will be normal during the next iteration.

Equipment

Figure 2 shows the Digitac computer. This computer is in a frame approximately 20 inches high, 26 inches wide, and 19 inches deep, a volume of about 5-1/2 cubic feet. All units of the computer are constructed on plug-in chassis, each chassis having a maximum of twelve tubes. The left half of the lower section contains the ten chassis of the arithmetic element and the control unit is in the remainder of the lower section. The wide chassis in the center of the rack contains the magnetic drum memory. The remaining chassis in the middle section are the memory read-write circuits and a group of timing pulse generators. The top portion of the rack contains the power wiring, heater transformers, and cooling system.

The entire computer contains 260 tubes and about 1300 diodes. The total power consumption is 1300 watts, including the cooling system.

Figure 3 shows the memory chassis when withdrawn from the rack. The drum is four inches in diameter, eight inches long, and rotates at about 7,000 rpm. It is a hollow aluminum cylinder with an oxide coating. There are 16 order heads, six number heads, two timing heads, and a pair of heads for a six-word circulating register. The pulse rate of the machine is 100 kc, which results in a recording density of about 75 cells to the inch. The non-return-to-zero form of recording is used. The read-back signal from these heads is a little over a volt peak-to-peak allowing the use of diode gating directly at the heads, rather than requiring individual preamplifiers before gating. Only one amplifier is used for each group of eight heads; the amplifiers are on the chassis to the left of the drum.

Computer Programing

The navigation problem the computer was required to solve in the air was essentially that of determining present position, selecting desired destination, and computing the control necessary to reach this destination. This over-all problem was, of course, actually composed of many smaller problems relating to either position determination, smoothing, or prediction, dependent upon the particular situation.

As previously mentioned, the basic position information is in the form of two time difference measurements. These measurements define two hyperboloids of rotation whose intersection is a curve. Knowing the aircraft's altitude it is possible to define a point on this curve of intersection, thus establishing the aircraft's position in three dimensional space. Figure 4 shows the general form of the position equations of the system; x, y, and z are present position and r is the slant range to the origin. At the time of turn-on of the computer, x, y, and r are assumed to be zero. Under these conditions, z is computed and then x, y, and r are computed in turn. Since each computation is based upon the other computed quantities, the first few computed quantities will not be accurate; however, it has been shown that this form of iterative solution will converge anywhere in the service area within six iterations. During normal computation the value of r is predicted from past values. This prediction allows the computation of accurate values of x, y, and z, even though only one iteration is performed each program cycle.

Figure 5 is a flow diagram of the navigation problem. In the initial portion of the routine a measurement of the two time differences is ordered. These measurements are then examined to determine if they fall within a reasonable range. If these measurements are unreasonable either due to interference or loss of signals, then the information is discarded and the lower branch of the routine is utilized. This dead reckoning branch utilizes the past computed positions, accumulated wind values, and aircraft instrument data to predict present position. If the measured time differences are usable, then the previously mentioned position solution is employed. In this branch of the routine the predicted or dead reckoned position is combined with the computed position to achieve a smoother position.

The final block of the normal routine is a computation of ground velocity, wind components, and a steering signal. Also at this time the remaining distance to the destination is checked. Assuming the destination has not been reached, then a return is made to the beginning of the routine. If the desired destination has been reached, either a change of destination is made or the final weapons computation is made.

It is of interest to note that the weapons computation does not normally require additional computing time. The dead reckoning branch of the routine requires considerably less time than the upper branch, and this time is utilized for the weapons computation. To insure that these computations are carried out periodically, the computer is forced into this routine every 32 cycles.

Figure 6 shows the method of position and velocity smoothing used. This smoothing is necessary since the computed positions will jitter due to noise on the received signals, the quantized nature of the time measurements, and the divergence of the hyperbolic lines. A predicted position is obtained from the past position and the distance traveled in one iteration time. The present position is obtained by combining the computed and the predicted positions. In the flight tests it was determined that the combination of these quantities should be equally weighted. Each cycle, a rough ground velocity is computed by the differences of position since the past iteration. The time base used in these velocity computations is not fixed but rather is a function of the routine length and the magnetic drum speed. To simplify the magnetic drum system this speed is allowed to vary and a portion of the time measurement equipment is used to measure the drum speed each iteration. The rough ground velocity is used in the cycle only to allow the computation of a rough wind. The smooth wind is accumulated by combining a very small portion of this rough wind with a major portion of the past wind value. A very heavy smoothing factor is used, about 31 to 1, assuming that the average wind value will not change rapidly. This heavy damping factor does not affect the prediction of position even during maneuvers, since the ground velocity used for prediction is formed by combining the smooth wind value and the air speed as presently measured from the aircraft instruments. All the computations mentioned are actually performed in rectangular coordinates requiring that the air speed be broken into its components based on the measured heading.

The aircraft instrument inputs to the computer are altitude, heading, and true air speed. The analogue-to-digital converters for these quantities are continuous reading devices that are sampled by direct order of the computer. Each instrument input quantity is checked by the computer to determine if it is reasonable prior to its use. If any quantity is not reasonable, the previous value is utilized.

The normal routine requires about half a second, during which time 360 operations are carried out. A total of 700 orders are required to include all the possible alternate routines and the weapon computations. Approximately 90 of the operations are multiplications or divisions.

In programing a problem of this type for a machine with a magnetic drum memory it is important to minimize the average memory access time. The access time is the time wasted waiting to obtain numbers or orders from the memory. In the normal cycle of this problem there are 660 references to the memory with an average access time of only one and one-half word times.

It is of interest to note that the program for this system uses the decision or alternate routine order more than 50 times. The extensive use of this type of order is one feature that makes airborne digital computation so attractive. By properly programing these decisions the computer is used to check all input data and select the proper method of computation to best utilize this information. Each input quantity, such as altitude, is first checked to determine if it is within a maximum and minimum bound and then is checked by comparison with the past value. If the quantity either exceeds the bounds or the change from the past value is too great, then this quantity is discarded and the previous value is used. This checking technique greatly reduces the effects of any noise present in the input data.

Conclusions

The Digitac System has very successfully shown that digital computers can be extremely powerful tools when applied to airborne control problems. The accuracy potential of digital computation is quite well known; it is not as well known that the versatility of digital computers can be their greatest asset. One computer such as the Digitac computer can be used as the computation element of many different systems without the need for any equipment changes. New situations in field operations can be handled by easily made program changes. Digital computers are definitely the airborne control elements of the future.

---

[1] W. L. Exner and A. D. Scarbrough, "A Digital Autopilot Coupler".

[2] E. M. Grabbe, D. W. Burbeck, and S. B. Neister, "Flight Testing of an Airborne Digital Computer".

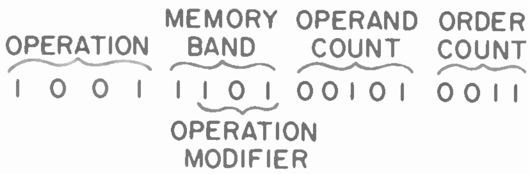


Fig. 1 - Order word.

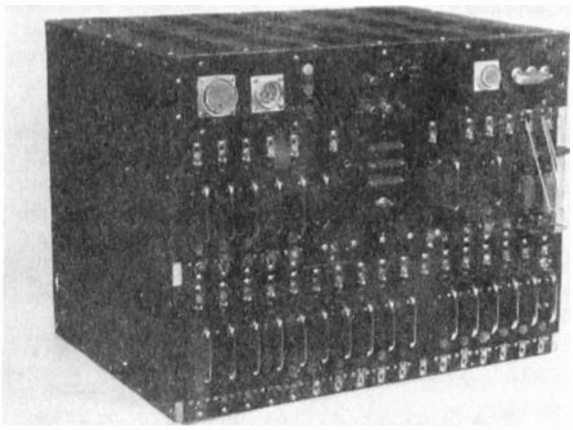


Fig. 2 - Digitac computer.

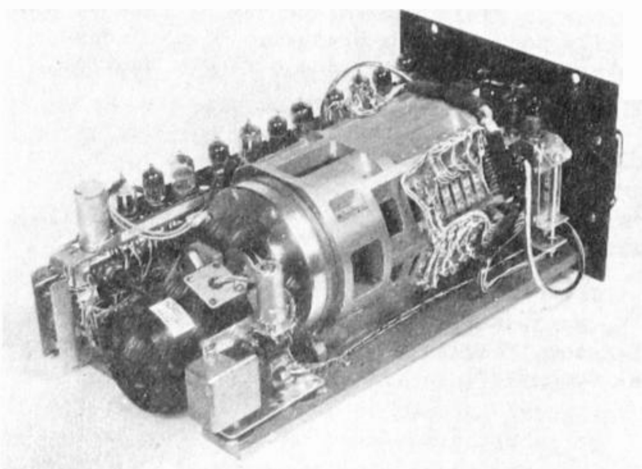


Fig. 3 - Memory unit.

$$X = f_1 (T_1, T_2, Z, R)$$

$$Y = f_2 (T_1, T_2, Z, R)$$

$$Z = f_3 (L, X, Y)$$

$$R = \sqrt{X^2 + Y^2 + Z^2}$$

Fig. 4 - Position equations.

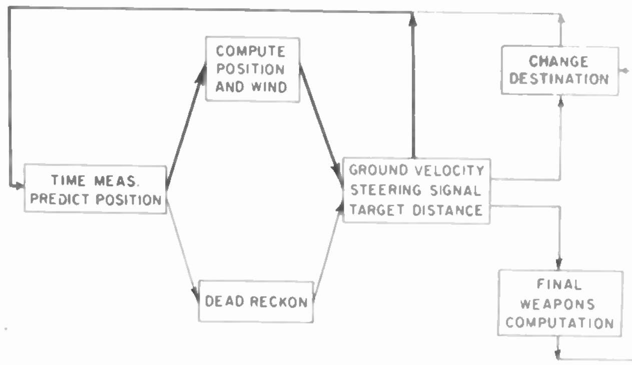


Fig. 5 - Problem diagram.

$\widetilde{P}$ = COMPUTED POSITION (rough)

$\overline{P} = P_{n-1} + (V_a + W_n)\Delta t$ (predicted)

$P_n = c\overline{P} + (1-c)\widetilde{P}$ (smoothed)

$\widetilde{V}_g = \dfrac{P_n - P_{n-1}}{\Delta t}$ (rough)

$\widetilde{W} = \widetilde{V}_g - V_a$ (rough)

$W_n = dW_{n-1} + (1-d)\widetilde{W}$ (smooth)

Fig. 6 - Smoothing equations.

A NEW FIXED-BEAM APPROACH SYSTEM

R. A. Hampshire
Federal Telecommunication Laboratories
a division of
International Telephone and Telegraph Corporation
Nutley, New Jersey

Summary--The paper describes a fixed-beam approach system for aircraft designed for the United States Air Force. The equipment is compatible with existing ILS installations and consists of a localizer operating between 108-112 megacycles and a glide path operating between 329-335 megacycles. The localizer incorporates the dual-beam system of operation.

A new design of Instrument Low Approach Equipment has been undertaken by the Federal Telecommunication Laboratories for the United States Air Force. This equipment works interchangeably with existing ILS equipment; the localizer being in the 108-112 megacycle band, the glide path in the 329-335 megacycle band, and each utilizing 90-150 cycle modulations. There have been added certain new features of interest. This is particularly true of the localizer which is of the dual-beam type. Investigations of this type of localizer were started at Wright Field several years ago. Similar investigations were conducted under the auspices of the ANDB at the Technical Development and Evaluation Center of the CAA at Indianapolis.

The principle of this dual-beam localizer is based on the proposition that very narrow lobes of energy radiated from the localizer antenna system will eliminate siting problems and that a superimposed localizer, operating on another frequency, will provide proper off-course indications outside the region served by the narrow main lobes. Parabolic reflectors, half-cheese antennas, broadside arrays, and waveguide antennas have been utilized for the generation of the narrow lobes. At the time we started our design, about three years ago, the choice of a best means to generate narrow lobes was difficult. We investigated parabolas and broadside antennas, and found that broadside antennas as developed by Andrew Alford Associates showed the most promise, and proceeded with our own specific design of a broadside array.

The basic design requirements may be explained by reference to Figure 1. The normal localizer provides left and right guidance to the airplane by modulating a signal with 90 cycles and 150 cycles in such a way that the 180 degree sector to one side of the runway centerline shows a predominance of 90 cycle modulation over 150 cycle modulation and the sector on the other side of the runway centerline shows 150 cycle modulation over 90 cycle modulation. The tone modulations are equal only on the runway centerline and the front and back extensions thereof. Now when an aircraft at Position A receives a signal not only from the localizer but also by reflection from some object such as the building shown, this reflected signal is characterized by a preponderance of one tone, and according to the r-f phase to which it adds to the direct signal, it will upset the equality of tone modulations one way or the other. This causes the aircraft's indicator to deflect to the left or right even though the aircraft is situated accurately on the runway centerline. As the airplane proceeds inbound towards its landing, the relative phase of the direct signal and the reflecting signal varies because the direct ray is being shortened more rapidly than the reflected ray. This causes the reflected signal to alternately add and subtract which causes the aircraft's indicator to swing back and forth--a phenomenon called "course bending." It may be seen that if the building were situated far to the side, this phase variation would be quite rapid. On the other hand, if the building were very close to the runway at Position B2, for example, the phase variation would be very slow. We can see in addition to this that the rapidity of phase variation is also a function of the distance out to the Point A where the aircraft receives the signal. As the aircraft flies in from a great distance, the bends become more rapid. This is a way of saying that the phase variation is more rapid as the angle between the reflected ray and the direct ray becomes greater. Now a reflecting object at Position B2, very close to the runway, will produce such a slow phase variation that the effect is a very minor shift of the course, free of abrupt bends and kinks. Since the type of reflecting object this close to the runway is small, the effect is not harmful. The building, however, being some 400 or 500 feet off the runway can produce harmful effects. A restriction of radiant energy to the sector between the dotted lines keeps signal off the building and is an effective means of reducing the course bends. As for reflecting objects at Positions B3 and B4 which may still lie within the localizer sector, these are too small and too remote and thus too little illuminated to cause trouble. Experience has shown that almost all objects large enough and close enough to the runway to cause trouble fall outside a sector some 12 or 15 degrees each side of the runway centerline.

Figure 2 shows the kind of radiation patterns utilized in the new localizer. The two narrow lobes define the localizer course. Note that the peaks of radiation are about three degrees off the course and the strength is reduced to zero at 12 degrees off course. The necessary course sharpness is produced with a crossover between the two lobes at values at about 80% of the lobe peaks. The ratios of on-course to off-course signal are such that

reflecting objects more than 8 or 10 degrees off course are not sufficiently illuminated to cause trouble even though they are quite close to the localizer.

Of course these radiation patterns have many minor lobes, which in this case are more than 20 db below the peaks of the main lobes. These minor lobes would cause false courses if only the narrow lobe radiation were utilized. There is therefore superimposed upon this basic localizer the second localizer having the broad radiation patterns shown. The on-course signal strength from the broad localizer is approximately one-third of the signal strength from the main localizer. The signal strengths become equal at 8 degrees off course and elsewhere, throughout the azimuth, the broad localizer, which we call a clearance localizer, predominates by a factor of at least three to one.

Now when these two localizer signals are radiated on carriers only 9 K. C. apart, and since the aircraft receiver has a much broader band than this, both signals are picked up. It is a well-known phenomenon that the modulation on the weaker signal is suppressed. This suppression of the modulation of the weaker signal is used to advantage here because any reflection of the clearance localizer signal which may be picked up by the aircraft while it is flying the course is suppressed and does not cause course bends nor oscillations of the aircraft's indicator.

Figure 3 is a diagram showing what happens in the receiver. On the left we see a situation where one signal is stronger than the other by a factor of three to one. Note that in this case the beat note is nearly a sinewave modulated by the tone of the weaker signal. Detection and averaging through the beat note yields the modulation envelope of the stronger signal. The phenomenon is described in Terman's text book and in a paper by Butterworth. On the right is shown a situation in which the two signals are equal. The beat note in this case is not a sinusoid. Sharp cusps are generated when the r-f vectors pass through the condition of phase opposition. Under these conditions, the result of detection and discarding the beat note does not restore the modulation due to the stronger signal alone. It is obvious from this that to achieve a smooth transition from a preponderance of one signal to the preponderance of the other--i.e., to maintain proper tone modulation--the modulation tones on both signals must be in the same audio phase and must modulate the carriers to the same depths. This is the condition shown in Figure 3.

Figure 4 shows a tracing which we have made of an actual flight recording across the localizer course. It has been traced accurately from the original Esterline-Angus paper. The smoothness of the crossover is apparent. The slight dips in crosspointer current are the effects of the transition which occurs approximately 8 degrees off course. The receiver output dips slightly here because the average value of the signal--that is, the value that effects the AVC--has been raised slightly by virtue of the beat note between the two signals not being sinusoidal. The effect is quite minor since the clearance is still well beyond the full scale indication of the aircraft meter and because the effect occurs about 8 degrees off course where monotonic indication is no longer required.

Figure 5 shows the equipment which generates these localizer signals. It is called Radio Set AN/MRN-7. The narrow radiation patterns are generated by broadside arrays of 12 dipoles mounted in front of a screen reflector. Horizontal wires on the stanchions eliminate the backward radiation. The array is 85 feet long and 7 feet high. The clearance signal is generated by an array of three Ramshorn antennas which is set up about 30 or 50 feet behind the main array. These antennas are a few inches higher than the upper edge of the main array reflector and provide for complete azimuth coverage without distortion by reflection or refraction from the screen. All of the transmitting equipment is located in the small van-type trailer shown in the background. For transport, the antennas and cables are stowed in the trailer and the antenna structure is collasped and carried on the roof.

Figure 6 is a diagram of the main array showing at the top the 12 dipoles arranged in a broadside array. Each of these dipoles is identical and is interchangeable. In order to generate the crossed lobes the current fed to the antennas is divided into two parts: one, the familiar dumbell, or reference, signal and the other, the sideband, or clover leaf, signal. Ten of the 12 elements are excited with carrier signal. The distribution of currents is plotted immediately below the antennas. The sideband signals fed to the array are distributed according to the curve superimposed. The sideband currents in one-half the array are, of course, in phase opposition to the other half and are each in quadrature with the carrier currents. Below is shown schematically the antenna distribution unit. Six regular hybrid bridge circuits are used. These prevent cross-feed between the carrier and sideband inputs and provide the proper phasing of antenna currents. The bridge on the right is not needed to prevent cross-feed but it is utlized to maintain phasing over the frequency band of 108 to 112 megacycles. The distribution of powers is arranged by taps on quarter-wave sections on transmission lines of adjustable ZO. These are so arranged that the standing-wave ratio on the lines is better than 1.1 over the band. All adjustments are made according to calculations of dimensions. No adjustment is necessary. The 12 cables which connect from the antenna tuning unit to the antennas are equal in length and are interchangeable.

The modulation at 90 and 150 cycles is generated by a mechanical modulator operated in principle like those now in use. The output of

a simple transmitter is divided into two branches and fed through two modulation troughs. Figure 7 is a schematic diagram of one of these troughs. It operates in a coaxial transmission line circuit and operates on the principle of inserting a tuneable wave trap in the outer conductor. The input and output lines project into the shielded box. The outer conductors terminate in a condenser, which is tuned and detuned by a motor-driven plate. The inner conductor is continuous. Since the device is symmetrical, the rotor plate is put in the neutral plane to prevent leakage of r-f energy along the rotor shaft. The equivalent circuit of the modulator trough is shown at the bottom of the figure. The untuned inductance is the section of bare inner conductor. The tuned inductor is the stub formed by the outside of these two outer conductors. The parameters are arranged so that the wave trap may be tuned towards resonance to achieve the modulation troughs and detuned in the capacitive direction for the modulation peaks. At the modulation peaks the fixed inductor nearly resonates the capacitive reactance of the variable section and provides for a flat line at the modulation peaks.

Figure 8 is a photograph of the transmitter modulator combination. All of the equipment is mounted on one front panel which hinges downward out of the cabinet. One side of the panel is occupied by the transmitter and modulator for the main array and on the other side of the panel by the transmitter and modulator for the clearance array. At the top of the panel are the two four-tube transmitters which generate the r-f signals; one for the main array and the other for the clearance array. Immediately below it, are the four modulation troughs situated side-by-side. One trough on each side is tuned by a five-blade rotor and the other by a three-blade rotor in order to achieve the 90 and 150 cycle modulations. A motor in the middle is directly coupled to the two shafts, thus providing accurate audio phasing. The hybrid bridge circuits which are used to separate the sidebands from the carrier are located between the two modulators.

Figure 9 is a photograph of this cabinet in operating position. The tuning controls for the two transmitters are shown at the top. The slide immediately beneath is an amplitude control unit which adjusts the amount of sideband power output so that the course width may be varied. It dumps unwanted sideband power into a dummy antenna. The adjustment is made without phase shift.

Figure 10 is a block diagram of the complete system. In one cabinet are located all of the power supplies and in another the transmitter and modulator which we have just seen. The outputs of the modulator go to a transfer unit and thence to the antennas. Note that with the power supplies are a voltage regulator and frequency converter. These permit operation on any line voltage between 95 and 135 or between 190 and 270 and on any line frequency between 45 and 75 cycles.

This transfer unit together with an alarm box and remote control unit are situated in another cabinet. The alarm box is the device which monitors the performance of the equipment. It is fed by three simple monitors set up in the field in front of the equipment at the locations proper to check course position, course width, and clearance. Complete standby equipment consisting of another transmitter modulator, another power supply, and a duplicate alarm box is provided. The equipment is turned on and off by remote control, it is self-monitoring and equipped for automatic changeover from main to standby.

The glide path equipment which provides for vertical guidance of the aircraft in the standard 329 to 335 megacycle band operates on the familiar null reference principle. This equipment is shown in Figure 11. The upper of the three antennas on the left is the null antenna which radiates a signal similar to the sideband, or clover leaf, signal of the localizer. The null in the vertical radiation pattern is, of course, produced by ground reflection. The antenna below this is the carrier, or reference, antenna situated at one-half the height above ground. Next is a three-element antenna radiating the same kind of signal as the upper antenna--that is, sideband only--this is the modifier antenna, and is used to straighten the glide path at its lowermost extremity.

Figure 12 is a photograph of the glide path transmitter modulator. It, like the localizer, is mounted on a single panel which hinges downward and outward from the cabinet for maintenance. The glide path, of course, has only one carrier frequency; thus, the cabinet contains only one transmitter and one modulator. The transmitter, at the top, utilizes five tubes. The modulator has two troughs which are in this case situated one on each side of a single motor. Immediately above the modulator is a small chassis containing tubes and crystals which are a portion of the built-in test equipment. Twenty kilocycle signals are generated here and fed to crystal detectors inserted in the transmission lines at the output of the modulator trough and in the carrier and sideband feed lines to the antenna system. These permit transfer of the modulation to a twenty kilocycle tone which is presented on a cathode ray oscilloscope in the equipment and provide for complete field maintenance adjustment of the modulators.

Figure 13 is a photograph taken inside the trailer of the glide path equipment. On the left is one transmitter modulator cabinet and on the right is the standby transmitter modulator. In the center there are, from top to bottom, the test oscilloscope, the main alarm box, the control unit, the standby alarm box, and the transfer unit. Dummy antennas are located in the transfer unit so that the standby or the main equipment may be passed into dummy loads for maintenance purposes while the main equipment is performing its normal functions. Two more cabinets, not shown, are situated in the other end of the

trailer and house the power supplies.

Although not specifically referred to previously, the localizer cabinet arrangement and test equipment facility is the same.

Figure 14 is a photograph of one of the three glide path monitors. The pick-up antenna is adjustable to provide for different glide path angles. As in the case of the localizer, three monitors are used: one for path position, one for path width, and one for path shape (modifier antenna radiation).

The two equipments, AN/MRN-7 and AN/MRN-8, use common construction and components wherever practicable. Every effort has been made to provide easy set-up and maintenance. The equipments are each self-contained, except for primary power which must be provided at either 115 or 230 volts; 4 K. W. is required. For transport, both antenna structures are carried on the trailer roofs and the cables, field monitors and antenna radiators are stowed inside the trailers. The number of vacuum tubes utilized in the power supplies and transmitters has been kept to a minimum. Reliable types of tubes have been used in the monitor, control and test circuits.

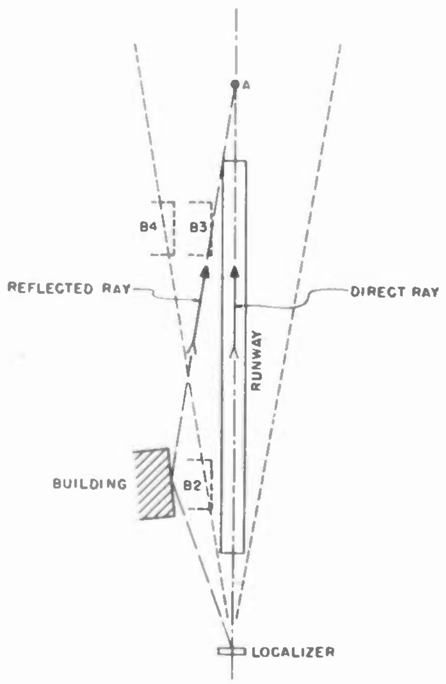


Fig. 1
Buildings may reflect radio waves to the aircraft at A in such a manner as to distort the localizer fields on which a safe approach depends.

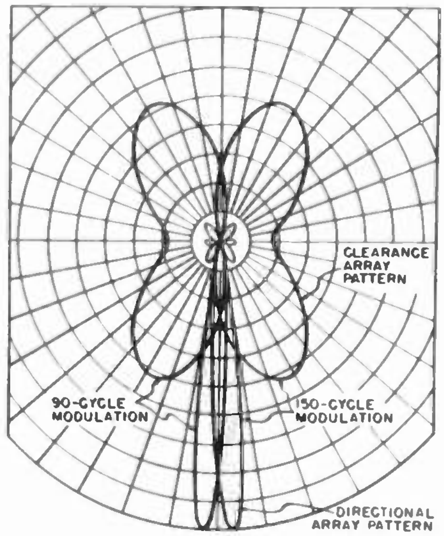


Fig. 2
Calculated directivity patterns of both the broad and narrow localizer radiations. The broad pattern permits the aircraft to find the narrow pattern that is used for the actual approach.

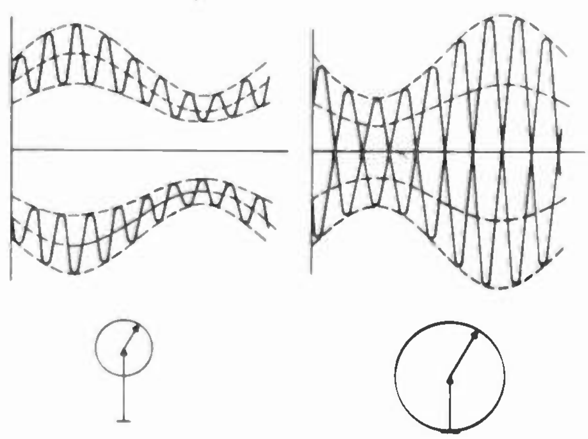


Fig. 3
At the left, one received signal is 3 times the intensity of the other while at the right both signals are equal in amplitude.

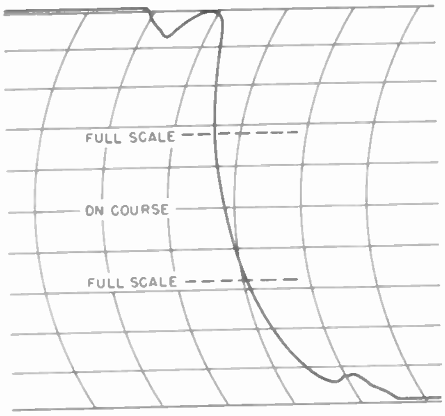


Fig. 4
Recording taken in a flight across the localizer course.

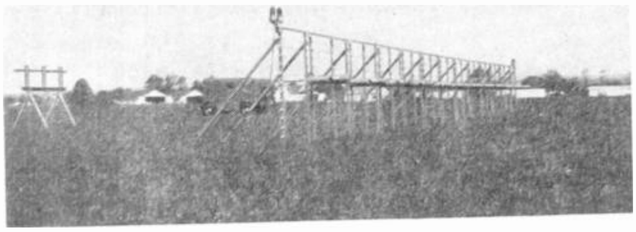


Fig. 5
The AN/MRN-7 equipment that produces both the broad and narrow localizer patterns.

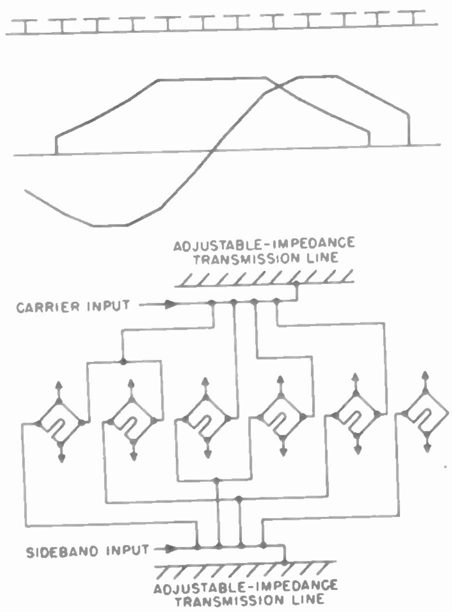


Fig. 6
Main array of dipoles with current distributions at top. Below is the method of supplying power to the antennas, which are connected to the vertical corners of the 6 radio-frequency bridges.

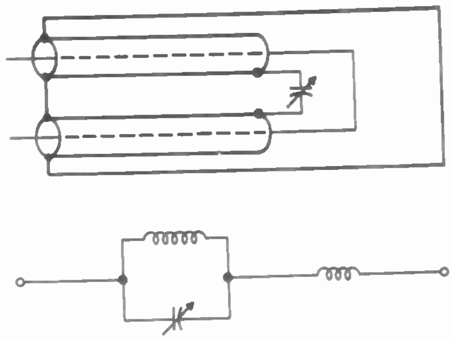


Fig. 7
Schematic diagram of modulation trough.

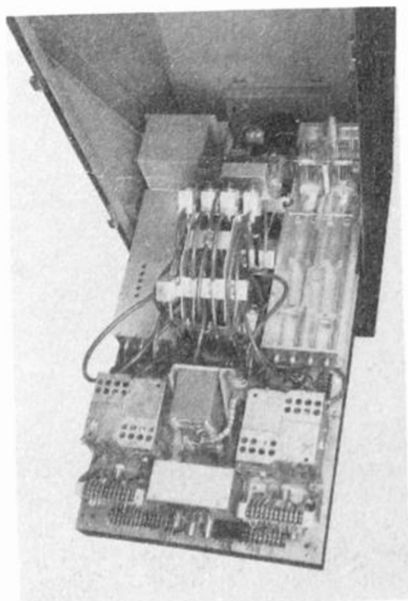


Fig. 8 - Transmitters and modulators.



Fig. 9 - Complete transmitter in its cabinet.

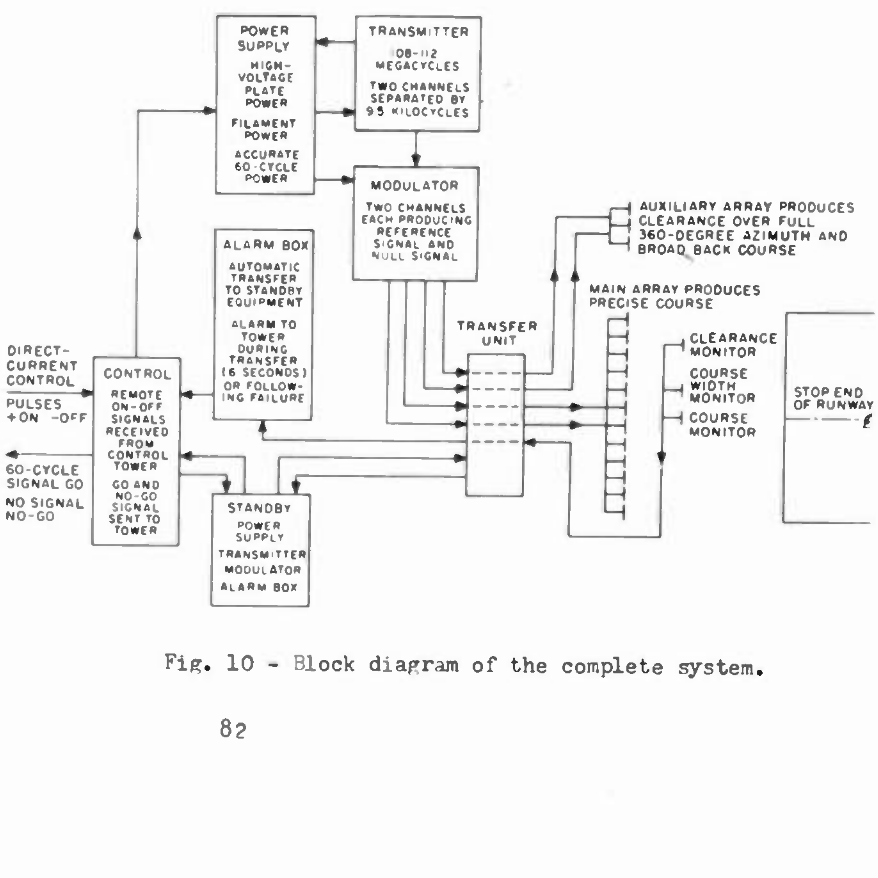


Fig. 10 - Block diagram of the complete system.

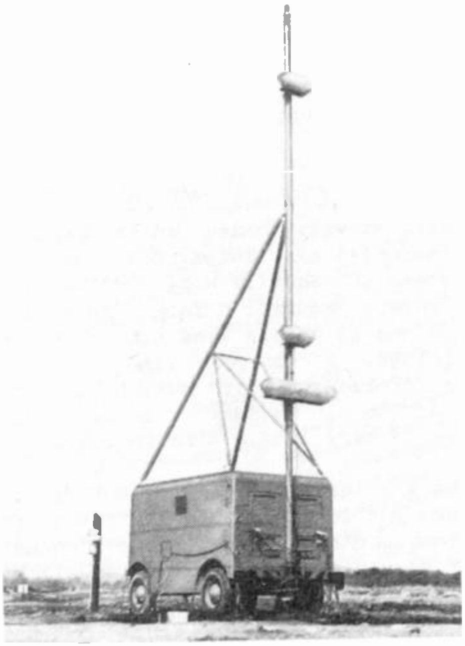


Fig. 11 - Glide-slope equipment.

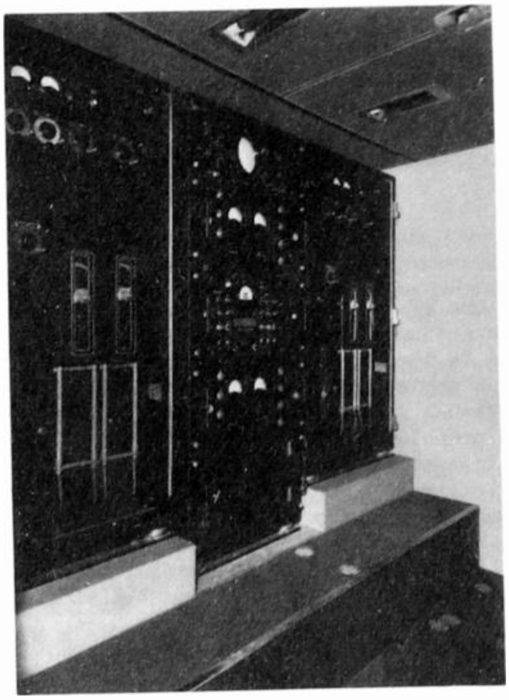


Fig. 13
Interior of the trailer containing the glide-slope equipment.

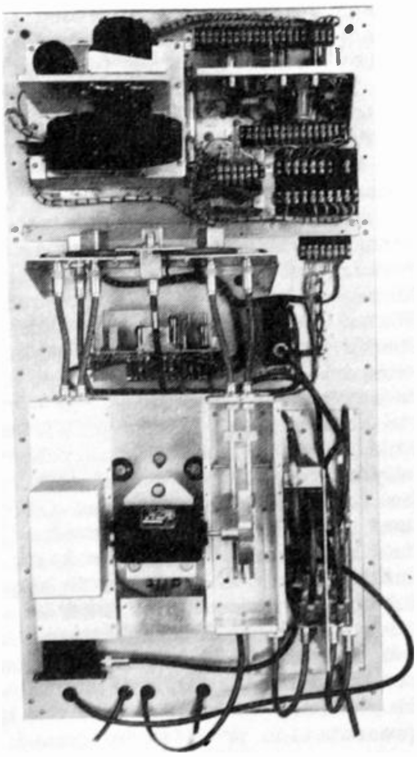


Fig. 12
Modulator for the glide-slope transmitter.

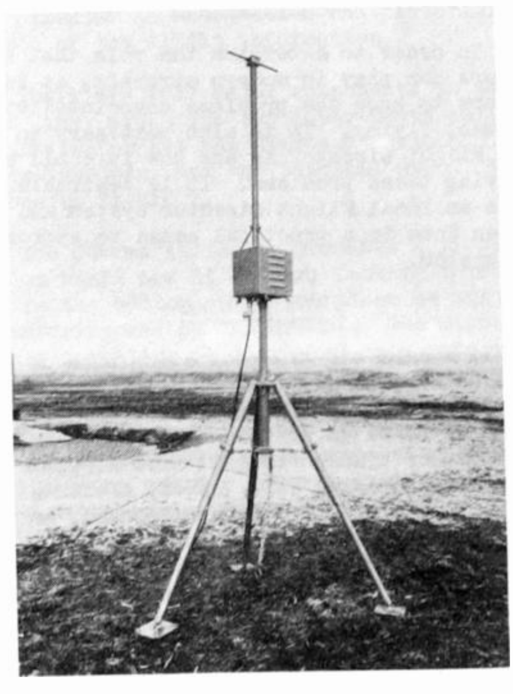


Fig. 14
One of the glide-slope monitors.

THE ROLE OF FLIGHT DIRECTORS IN PRESENT-DAY AIRCRAFT

N. L. Graham
Collins Radio Company
Cedar Rapids, Iowa

Summary

One of the more recent additions to the ever expanding list of aircraft navigation systems is the Flight Director. A Flight Director is a semiautomatic system which provides the pilot visual guidance information in accordance with a preselected flight plan. The information normally available on the flight panel is fed electrically into a computer where it is modified and combined. The resultant information is presented to the pilot visually in very simple form, thus eliminating most of the mental calculations usually associated with instrument flying.

The Collins Radio Company has spent many years developing such a Flight Director system. The outstanding feature of this system is integration and pictorial presentation of the navigation information. This had been done by introducing two new instruments, the Course Indicator and the Approach Horizon. Intensive flight tests conducted on the Collins System have demonstrated the high degree of accuracy which can be expected from Flight Directors.

Introduction

In order to determine the role that Flight Directors can play in modern aircraft, it is necessary to know the problems associated with instrument flying. It is also necessary to know what a Flight Director is and how it could help in solving these problems. It is desirable to discuss an ideal Flight Director system and what has been done in a practical sense to approach an ideal system.

The Problem

In modern air transportation much of the navigation is done with the pilot relying solely upon his instruments. This is necessary if the airplane is to be an efficient and reliable means of all-weather transportation. The airplane has been improved and its size, speed, altitude capabilities, and range have been increased, making it more nearly an all-weather machine. Many new types of equipment have been invented for instrument flying and instrument upon instrument has been added to the flight panel. This has been done by many different groups independent of each other and without due regard to the overall problem of instrumentation. The result is a crammed and complex instrument panel. The one thing in the aircraft that has not changed is the pilot. Fundamentally he is the same man that has been flying for many years. The result of the increased complexity of the machine is that the man becomes more heavily loaded and at times the demand has exceeded his limitations. With the increased speed and heavier wing loading, the modern aircraft requires a larger maneuvering radius and the pilot has less time in which to make decisions. At the same time he is provided with more information which must be analyzed before any action can be taken. If he makes a mistake there is very little time to correct it.

The most critical part of any flight is the descent and landing. The transition from flying instruments to visual ground contact and touchdown places a heavy load on the pilot. Even with the aid of such systems as GCA, Automatic GCA, and ILS, a low approach places the pilot under great strain. It is during this part of the flight that the pilot is most in need of equipment to reduce his load.

During an approach the pilot flies his airplane down a path framed by the intersection of two imaginary planes, one in the horizontal axis and one in the vertical axis. The vertical plane extends upward from an extension of the runway. The horizontal plane extends from the point where final approach starts to the intended point of ground contact. To fly a smooth approach the pilot must have three types of information for each axis. To determine his position he must know the lateral and vertical displacement from the path. To visualize how to fly and maintain the path he must know his heading and pitch and rate of change of these quantities.

During and ILS instrument approach the localizer radio beam provides information in the horizontal plane and the glide slope radio beam provides it in the vertical plane. The heading information is furnished by a gyro stabilized magnetic compass or a directional gyro and the aircraft's attitude is obtained from a vertical gyro. With conventional flight panels this information is presented on many instruments in a wide variety of forms. The pilot must scan these instruments to receive the required information. He must then form a mental picture of where he is, where he has been, and where he is going. From this he can determine what action is required to achieve his goal. This process must be carried on continuously without interruption if he is to fly a smooth approach. When visual contact is made he must transfer the source of his information from the many instruments to the more familiar presentation provided by ground objects.

The information presented on the conventional flight panel is in a form which differs widely from that provided the pilot during visual

flight. For this reason it is necessary for a pilot to go through a completely new learning process to fly by instruments. In fact, it requires almost the same amount of time to train a pilot to fly by conventional instrumentation as it does a non-pilot.

General Solution

There are two possible solutions to the problem of reducing the load placed on the pilot during a final approach. One is an automatic system and the other is a semiautomatic system. A Flight Director is an example of a semiautomatic system. A Flight Director, as specified by the Society of Automotive Engineers, is a device which indicates to the pilot, by visual means, the correct control application for the operation of an aircraft in accordance with a preselected flight plan. The displacement, heading and attitude information normally available on the flight panel is fed electrically into the Flight Director, where it is modified and combined as the situation may demand. The resultant information is presented visually in very simple form. All the computing has been done and the pilot has merely to maneuver the plane so as to satisfy the Flight Director indicator. In a semiautomatic system such as this, the pilot is actually being used as a servo actuator to close the control loop. Flight Director systems which fulfill these basic requirements are currently being produced by several companies.

Ideal Solution

So far, the basic principle of Flight Directors have been discussed. We should not stop here but go on to the problem of overall instrumentation. The Flight Director should not be one more system to be squeezed into the instrument panel.

The Flight Director information may be presented in a variety of ways. The most desirable way would be to present it in a manner the pilot is accustomed to seeing in his everyday operation. It would be desirable to make the instrumentation appear to be a duplication of what he would see out the cockpit window. He should see the horizon and his attitude with respect to it. It should have the same sensing and sensitivity that he is accustomed to seeing during contact flying. He should see at a glance the vertical and horizontal displacement from his selected track and the manner in which he is approaching it. The picture should present all the information the pilot needs without the necessity of referring to numbers. Numbers can be misread and misinterpreted during moments of high stress. The instrument presentation should provide accurate and concise information at a glance so that this could not happen.

Another desirable feature of an ideal Flight Director system would be integration of the instrumentation. We should not just add a picture to the flight panel but should use the information that is already present to form the picture. Certain basic flight information could not and should not be removed from the instrument panel. This information should be presented, as nearly as possible, in its conventional from and yet be integrated with other information to give the overall picture of the operation.

A Flight Director should provide the same pictorial guidance information during enroute navigation. It should provide all the necessary information to fly VHF omnirange and magnetic heading. It should enable the pilot to see a dynamic display during standard maneuvers such as procedure turns, holding patterns, and the like. Simplicity and reliability should certainly be a prime consideration in the design of a Flight Director system. The system should be simple in operation and be as reliable as present-day design permits. In this respect the number of tubes and similar components should be held to a minimum or eliminated. The simplicity of operation should be stressed and the number of controls should also be held to a minimum. Such features as automatic cross wind correction should be incorporated to prevent the necessity of changing of any of the Flight Director controls during final approach.

Practical Solution

As an example of what can be done along these lines, the Collins Radio Company has spent many years developing a Flight Director system embodying as many of these desirable features as possible. The outstanding feature of the Collins Flight Director is integration and pictorial presentation of the flight information.

There are four major components in this system. They are: The two instruments called the Course Indicator and the Approach Horizon, the Steering Computer, and the Vertical Gyro. These are shown in figure 1.

The Course Indicator presents a picture the pilot would see if he were looking through a window in the bottom of the airplane at his radio track superimposed on the ground. The magnetic compass information is displayed on a rotating card. The aircraft heading is always read under the lubber line at the top of the instrument. A heading selector knob located on the lower left of the instrument provides a means of marking the desired heading and for supplying the steering computer with heading error information. The lower right hand knob provides means of omni-bearing track selection and the bar in the center provides deviation indication. To-from indication is provided by the arrow in the center of the instrument. To complete the picture a symbol of the airplane is etched on the face of the instrument.

This instrument provides integration of the magnetic compass repeater, omni-bearing selector, and lateral displacement from the deviation indicator. The situation of the aircraft can be seen at a glance without the necessity of referring to numbers. The plane shown in figure 2 is

flying a magnetic heading of nearly due North. The bearing of the desired track is approximately 330 degrees. The pilot is flying toward his course and is approaching it at an angle of approximately 30 degrees. He is still 3 dots off course. The To-From arrow indicates he is flying to the station.

The Approach Horizon presents a picture of the aircraft's position in the vertical plane such as the pilot would see out the cockpit window plus the computed lateral guidance information. The small bar at the left is glide slope displacement information. The aircraft's attitude is presented by the pitch bar and the horizon bar. These two indicators function similarly to a conventional horizon. The vertical pointer is the Flight Director steering indicator which presents the computed information to the pilot. Warning flags are provided for localizer and glide slope receiver failure.

A pitch trim control is provided at the lower left corner of the instrument which functions in its conventional capacity to adjust the pitch bar for level flight during enroute navigation. The pitch attitude for approach is fixed and depends upon the type of aircraft. The control at the lower right is for selecting mode of operation. One position is for enroute navigation and the other for final approach.

This instrument presents integrated information from the attitude gyro, vertical displacement and warning flags from the deviation indicator, and Flight Director steering information and function selection. When flying a computed signal a pilot has his aircraft's attitude information directly in view without reference to other instruments. In figure 3 the aircraft is above the glide slope and the pitch of the airplane is down, with respect to the normal approach attitude, so as to intercept the beam. The steering pointer indicates the pilot should bank to the right and the horizon bar indicates he is now banking about 8° right.

In integrating the flight information the basic navigation information is not removed. The course bar and the glide slope pointer are basic displacement information direct from their respective receivers. The compass card is a servoed compass repeater which repeats the information supplied by a stabilized magnetic compass. The pitch and bank information is supplied by the remote indicating vertical gyro.

Simplicity and reliability were definite goals. There are only two operating positions, one for enroute navigation and one for approach. The complete system contains only six tubes and incorporates an automatic cross wind correction circuit when in the approach position.

The accuracy of a Flight Director system is of a high order. Tests were conducted on the Collins Flight Director at ten different airports. Ten approaches were made at each of the airports and the results were recorded. The composite data is presented in figure 4. It shows that under the widely varying conditions that existed, 50% of approaches were within 15 feet of the localizer beam centerline at the end of the runway. Ninety percent of the approaches were within 40 feet and 100% were within 95 feet. Since most major airports have runways 200 feet wide, all approaches could be considered as satisfactory.

Similar tests on a more detailed and precise basis conducted at the Air Development Center, Wright Patterson Air Force Base, Dayton, Ohio, con firmed our results.

Conclusions

The pilot is in need of assistance in flying modern aircraft. This is especially true during an instrument approach. The Flight Director, a semiautomatic system, furnishes the pilot simple guidance information during an instrument approach and enroute navigation. The ideal system would present the information at a glance in a concise and accurate manner. The Collins Flight System strives to accomplish this by a pictorial presentation. Tests conducted on Flight Director systems have shown them to be accurate and reliable. It appears that the Flight Director has a definite role to play in present and future aircraft.

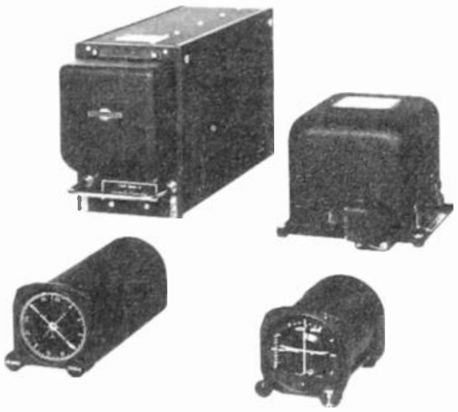


Fig. 1
Components of the Collins Flight Director System.



Fig. 3 - Approach horizon.



Fig. 2 - Course indicator.

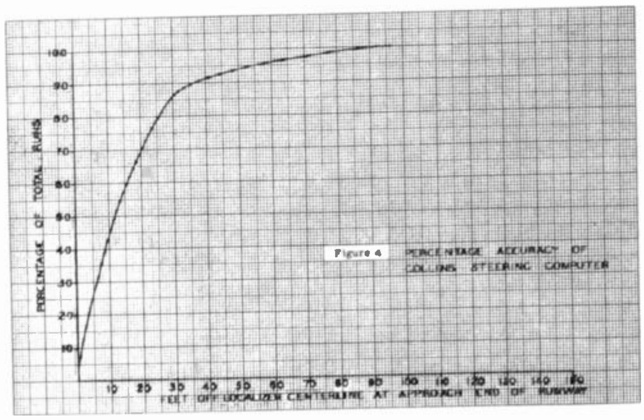


Fig. 4
Accuracy of the Collins Flight Director System.

NAVAGLOBE — NAVARHO LONG-RANGE RADIO NAVIGATIONAL SYSTEM

C.T. Clark, R.I. Colin, M. Dishal,
Federal Telecommunications Laboratories, a division of
International Telephone and Telegraph Corporation; Nutley, New Jersey

I. Gordy,
Rome Air Development Center; Griffiss Air Force Base, New York

and M. Rogoff
Now With the McDermott Company; Trenton, New Jersey

Navaglobe is a long-distance radio navigational system under development by Federal Telecommunications Laboratories for the United States Air Force. For obtaining reliable long-range service, it operates in the low-frequency band and with very narrow bandwidth. Service is omnidirectional and the airborne bearing indications are automatic. The principles of the transmitting and receiving equipment are described, including features designed for minimizing the effects of atmospheric noise. Views of actual equipment are shown, and the various stages through which the developmental program has progressed are discussed. Performance of the experimental equipment is described, including the results of transcontinental and transatlantic flight tests. Work and future plans for the incorporation of distance measurment by means of phase comparison of the Navaglobe signals with reference to an airborne crystal-controlled frequency generator, forming the complete Navaglobe — Navarho system, are discussed.

• • •

The Navaglobe project has been carried on for the United States Air Force by Federal Telecommunications Laboratories, with the goal of providing a radio navigational system that is truly long-range and reliable, and that gives automatic readings. Construction of experimental transmitting and receiving equipment has progressed to a stage that has made possible extensive laboratory, field, and flight tests. Performance in these tests, indicating that the system will fulfill its goals, makes it timely to present a review of this project that may result in a valuable new facility for this age of global air operations.

Long-range air navigation today is done essentially by dead reckoning, as it was 10 and 20 years ago[1]. Because of the cumulative effect of inaccuracies in estimates of distance and direction, dead reckoning on long flights should be checked as frequently as possible by positive position determinations; the goal, of course, is a supplanting system that could provide reliable position determination at will or continuously. To obtain position determinations by celestial methods requires complicated time-consuming calculations based on observations by trained navigators in clear weather at certain times of day and on a well-stabilized aircraft. These conditions are unduly restrictive in air operations, whether civil or military. As regards position determination by radio methods, there are certain facilities available for this purpose from time to time and place to place, but there is no single radio aid in use that does give or could consistently give navigational guidance during the entire distance of transoceanic or transpolar flights.

For air traffic over populated land areas, where there is no problem in locating a sufficient number of ground stations, short- and medium-range radio aids of various types successfully provide continuous navigational guidance to contribute to the safety, efficiency, and economy of air operations.[2] In other regions that constitute the larger part of the earth's area, the basic problems involve geography (availability of land sites suitable for the location of navigational transmitters) and radio propagation (the range required of the transmissions if they are to reach aircraft flying anywhere over oceanic, polar, and other nonpopulated regions with the reliability that is vital for a navigational system). Some existing radio navigational systems approach a satisfactory range, but only under certain limited propagation conditions; at other times their range falls quite short of the needs, so that they are really quasi-long-distance systems. None of these automatically give a direct reading and so are unsuitable for cockpit use by the pilot, nor are they omnidirectional in coverage.

1. Basic Reliability Considerations

The Navaglobe project started with a study of the basic problem mentioned above[3]. Examination of the globe showed

that there are geographically available sufficient and suitably related land sites for navigational transmitters to cover, ultimately, all oceanic and polar regions with radio position-fixing service, provided that the reliable day-in-and-day-out range of the transmitters is around 1500 nautical miles (2780 kilometers). The next step was to determine the general parameters of a radio system to fulfill that requirement. A detailed survey of data on radio transmission in all portions of the radio spectrum was undertaken, considering also the related practical factors of atmospheric noise intensities and antenna radiation efficiencies. Statistical studies were made on all these points of data accumulated over many years by government agencies and commercial communication companies.

The conclusion, since supported by independent studies, was that the practical hope for reliable long-range radio navigation lay in low-frequency operation, in the vicinity of 70 to 100 kilocycles per second. The International Telecommunications Convention has since reserved the band between 90 and 110 kilocycles for this type of service[4]. At the same time, very-narrow bandwidths of the order of 20 to 100 cycles per second should be used. This is not only to conserve channel space, always a critical practical problem at low frequencies, but also to produce useable signal-to-noise ratios at long distances without requiring prohibitively expensive transmitter powers and antenna sizes.

The original Navaglobe proposal,[5,6] and the equipment since developed and tested, conforms to the conclusions of the basic study, for it was recognized that propagational reliability is the critical element of a long-distance radio navigational system, to which must be added the usual requirements of equipment reliability. The Navaglobe transmission rate of 4 continuous-wave signals per second is compatible with very-narrow bandwidth, low-frequency operation, and yet is a fast-enough information rate for long-range navigational purposes. A major effort went into designing the system and equipment so as to minimize the effects of atmospheric noise generated by local thunderstorms, which is particularly serious in midsummer and near the tropics. Such noise is generally the limiting factor in obtaining consistent long-distance operation of low-frequency radio circuits. As a further measure to insure reliability and to instill confidence in the pilot, the system provides automatic fail-safe indications and automatic calibration-checking indications. A description of these features and of the general principles of operation of the Navaglobe system follows.

2. Principles of Operation

The Navaglobe ground installation consists of three antennas situated at the corners of an equilateral triangle, as indicated in Figure 1. The antennas are spaced 0.4 wavelength apart, which amounts to approximately 3600 feet (1100 meters) when transmitting at 100 kilocycles. Radio-frequency power is supplied to the antennas according to this schedule: currents of equal strength and phase go to antenna pair 1-2: then similarly to antenna pair 2-3; and finally to antenna pair 3-1. This cycle is constantly repeated at a rate of 1 per second. The timing and switching apparatus for controlling this cycle is located in a central hut, which also houses equipment for monitoring the phase and amplitude of the currents in the three antennas.

Antenna pair 1-2 produces a radio signal with a strength greatest in the directions perpendicular to the line joining the antennas and weakest in directions parallel to that line. The systematic variation in strength along different directions is indicated graphically to scale by radiation pattern AA in Figure 1, which resembles the figure 8. Patterns BB and CC, produced, respectively, by antenna pairs 2-3 and 3-1 are identical in shape but their axes are shifted 120 and 240 degrees from that of pattern AA.

The net result is that three signals, A, B, and C, are radiated over and over again, and that the relative strengths of the three signals are different along each direction. At bearing 40 degrees, for example, the relative strengths of the three signals received in time sequence are as shown to scale in the upper part of Figure 2. At the receiving end, the three signals are isolated and applied to a bearing translator circuit that automatically rotates a shaft to a position that depends on the relative strength of these signals. Thus a pointer attached to this shaft indicates the bearing of the receiver from the transmitter. From the nature of the overlapping radiation patterns and the method in which the information in the resulting signals is utilized, it is evident that the Navaglobe station provides omnidirectional bearing-indication service.

The signal denoted S in Figure 2 is the synchronizing signal. It is radiated once each cycle from a single antenna and hence is received with equal strength in all directions. It is radiated on a

slightly different frequency from that used for the A, B, and C signals, and thus is separately indentifiable in the airborne receiver. The S signal marks the start of each transmission cycle and enables the receiving equipment to isolate and identify the A, B, and C signals for proper application to the bearing-translator circuits.

The elementary principle of the bearing translator is explained in the upper half of Figure 3. The A, B, and C signals from the receiver are fed to the designated stator inductors through gating and switching circuits controlled by the synchronizing signal. Coming from the receiver in direct-current form, these signal currents produce successive individual direct-current magnetic-field components aligned with the respective inductors. The angle Ø of the resultant magnetic field is assumed by the magnetic needle and depends on the relative strengths of the A, B, and C signals, and hence on the bearing from the transmitter array.

Elementary indicators of the above-described "ratiometer" type were used in the initial experimental demonstrations of the Navaglobe principle. For this purpose a three-antenna transmitting installation was erected and placed in operation near Adamston on the New Jersey coast. This station, however, was built for operation at 150 kilocycles since a channel in the preferred 100-kilocycle band was not obtainable at the time. Exploratory field tests of system operation with this transmitter were performed in 1951, the receiving equipment being installed in a motor vehicle that travelled to points at various distances and directions. While reception was effected at times as far as Colorado Springs, Colorado, and the indications were in close accord with the theoretical calculations, the need for certain improvements was indicated, particularly in regard to the system behavior under heavy local-thunderstorm atmospheric-noise conditions. Improvements were accordingly made in the bearing translator, receiver, and synchronizing system.

The principle of the mathematically equivalent but practically superior "resolver" form of bearing translator now used is also shown in Figure 3. Here the A, B, and C signal voltages from the receiver's intermediate-frequency output pass in alternating-current form to the designated stator inductors under control of the synchronizing system. In the rotor or pick-up coil, individual voltages are successively induced that depend on the strengths of the A, B, and C signals and also on the angular position of the rotor coil. An integrating-type square-law detector (essentially a watt-hour-meter device) measures the algebraic sum of these induced voltages over an integral number of transmission cycles, and its output serves as an error signal to control a servomechanism. The servo automatically positions the rotor coil to an angle Ø at which there is a null output from the integrating detector. This angle Ø, as before, depends on the strength ratio of the original A, B, and C signals and hence on the bearing of the transmitter array.

To insure that the signals going to the bearing translator and detector are at the optimum level for distortionless operation of these devices, an automatic-gain-control circuit is provided. Although this circuit acts relatively quickly in correcting the gain of the receiver as required, it is so devised as to do this without changing the A, B, and C signal strengths relative to each other. A flag alarm warns against reading the bearing indicator before the automatic-gain-control system and the automatic synchronizing system have had time to bring about proper operating conditions, or when they are prevented from doing so because of extremely weak, noisy, or otherwise unsatisfactory signals.

As remarked earlier, atmospheric noise is generally the limiting factor in obtaining consistent operation of long-distance low-frequency radio circuits. In Navaglobe operation, such noise piled on top of the signal could mask the proper strength ratios of the A, B, and C signals and so result in erratic bearing readings. Noise effects are minimized by use of square-law detection, very-narrow-bandwidth operation, and by signal-amplitude limiting in the initial wider-band stages of the receiver. The effectiveness of these measures in laboratory tests is demonstrated graphically by the oscillograms shown in Figure 4. Flight tests through actual thunderstorm areas have equally demonstrated the antinoise effectiveness of the improved receiver. Under the most severe conditions, useable bearing indications are still obtainable by increasing the reading time. For this purpose, the bearing translator is switched to a mode of operation wherein, in effect, it examines a large number of cycles of A, B, and C signals before it gives its answer; this tends to average out the disturbances caused by noise in one or a small number of cycles.

A second possible effect of atmospheric noise is to obliterate the synchronizing signal and hence incapacitate

the bearing translator even though there may still be useful information in the A, B, and C signals. To minimize this effect, the airborne equipment contains a tuning-fork oscillator that is automatically kept locked in step with the recurring synchronizing signals received from the ground station. Should the latter signals be interferred with because of noise, the inherent frequency stability of the tuning-fork oscillator prevents the airborne timing cycle from being driven very far out of phase with the ground transmission cycle. Ordinarily, the ground synchronizing signal may be disturbed seriously for only a few seconds at a time; the airborne fork has proved to be stable enough to maintain adequate synchronization for lapses up to nearly a half-hour.

The automatic behavior of the receiver-indicator circuits with respect to noise may be likened to the mental action of a person viewing a noisy signal on an oscilloscope; by recourse to averaging and memory processes, the maximum information available from such a signal is extracted. In the extreme, should the signals be so bad as to contain no trustworthy bearing information, the automatic flag alarm warns the pilot, as it does in cases of equipment malfunctioning.

As a further safeguard and reassurance to the pilot, it is proposed that each Navaglobe transmitter will send out on the hour a special signal pattern that is identical in all directions. This pattern will correspond to a special check bearing, say 55 degrees, known in advance to all pilots. Regardless of where he is, every pilot should observe a 55-degree reading on his Navaglobe indicator during this checking period. Thus the pilot is periodically given simple and effective assurance that the airborne equipment is properly operative and calibrated.

3. Test-Flight Performance

Equipment operating according to the principles described above was used in a program of flight tests performed during the summer of 1952. A view of the experimental type of airborne equipment used in these tests is given in Figure 5. This equipment is bulky; the large indicators were specially designed for testing, calibrating, and recording purposes. (Descriptions of the type of equipment operating with the same basic circuits but packaged more compactly for operational use aboard aircraft are given in the next section; see also Figure 10.)

Consistent service out to very-long distances being the indispensable requirement for a long-range navigational aid, experimental determination of the performance of Navaglobe in this respect was the primary objective of the flights. The summer season was chosen because midsummer daytime conditions provide the severest test for long-distance operation of low-frequency radio circuits, and low-frequency operation is conceded to offer the best hope for reliable long-distance radio navigation. The Navaglobe equipment was installed in an Air Force Type C-54 aircraft, accompanied by engineers and observers from the Federal Telecommunication Laboratories and from the Rome and Wright Air Development Centers of the United States Air Force.

For these tests, it was decided to use 100-kilocycle transmissions rather than 150-kilocycle transmissions, since all indications are that 100-kilocycle propagation is substantially better at 1500-mile (2780-kilometer) ranges. This precluded using the Navaglobe station at Adamston, which is restricted to 150-kilocycle operation by the physical spacing of its three antenna towers. A 100-kilocycle transmitter and efficient single-tower radiator were available at Forrestport in central New York State. This nondirectional station was used to radiate a selection of signal patterns simulating the cycle of signals that would be produced along certain specific directions by a Navaglobe 3-antenna array. The particular pattern was changed every hour in a random sequence not known in advance on board the aircraft. During a 10-minute interval of every hour, a steady continuous-wave signal was radiated to permit field-strength measurements to be made.

The flight program included an overland phase from New York to California and return and an oversea phase, via Bermuda, the Azores, England, and Iceland. In addition to the round-the-clock visual inspection of the bearing indicators, oscillographic monitoring of the incoming signals, and periodic field-strength measurements, continuous pen recordings of the bearing indications were also made over some 12 days of testing, partly aloft and partly on the ground.

A chart summary of the flights is shown in Figures 6 and 7. Satisfactory reception of signals and operation of the bearing indicators was maintained in the daytime out to beyond Prescott, Arizona, and out to a point halfway between the southern tips of Iceland and Greenland. These points are approximately 1800 nautical miles (3330 kilometers) distant

from the transmitter in New York state. Nighttime operation was observed out to San Bernardino, California, the limit of the transcontinental route, some 2000 nautical miles (3700 kilometers); and on the oversea flight to a point a few hundred miles southwest of England at which time daybreak occurred. The latter case represents a range of approximately 2600 nautical miles (4820 kilometers).

On one portion of the overland flight, north of El Paso, Texas, a violent summer-thunderstorm area with its high atmospheric-noise level was passed; the incoming signal deteriorated and the bearing indication fluctuated somewhat, but useable indications were still produced. Near Greenland on the transoceanic flight, the aircraft flew through a precipitation-static area. All other radio facilities aboard were inoperative for over a half hour, but the Navaglobe signals produced useable bearing indications except for about 7 minutes.

The true nighttime limit is not known, for although the European flight continued farther, signals from some powerful European low-frequency transmitters caused interference while the aircraft was near the European mainland. Also, it was found out in flight that some conducted noise from a local power source on the airplane had been deteriorating performance. These facts indicate that the maximum day and night ranges observed might be bettered.

Undoubtedly winter performance and ranges would be considerably better, both day and night, since atmospheric noise levels are substantially lower in that season. However, the practical gauge of a radio navigational system is its performance under the least-favorable conditions, in this case midsummer daytime. A range of 1800 nautical miles (approximately 2100 land miles or 3330 kilometers) under such circumstances, overland as well as oversea, is unparalleled for radio navigational systems, even those of the nonautomatic-reading type. Even discounting the probability that still better ranges may be achieved with higher transmitter powers and with equipment improvements suggested by experience during the flight tests, the demonstration of 1800-nautical-mile (3330-kilometer) midsummer daytime operation by the experimental Navaglobe equipment indicates that the system will fulfill with comfortable safety margin the prime requirement for a truly long-distance radio navigational facility, that is, sufficient range to make possible reliable position-fixing service over all oceanic and polar regions.

4. Airborne Presentation and Installation

This section describes prototype airborne equipment that operates with the basic circuits and designs proved in the flight tests, but which is packaged more compactly and built to conform to standard Air Force specifications for airborne electronic equipment. While reliability is the primary consideration for a navigational aid, the size of the installation, and especially the simplicity of its use by the pilot, are important in determining the usefulness of any airborne equipment.

In Navaglobe, the numerical indications required for position fixing or for following a straight course to or from a selected ground station are presented in the simple terms familiar to pilots who navigate overland by conventional short-distance radio aids, such as the radio compass (ADF) or the omnirange beacon (VOR). That is, the automatic Navaglobe meter directly indicates in degrees the bearing of the observer from the selected ground station.

In the interest of standardization and economy of cockpit installation, Navaglobe bearings may be presented on the standard meters used for radio-compass, omnirange beacon, or instrument-approach (ILS) services; namely the ID-249 and the ID-250 meters. The Navaglobe airborne equipment has outputs provided to operate either or both of these standard indicators; but these instruments cannot be read closer than within one or two degrees.

Figure 8 shows the special Navaglobe indicator developed for cockpit or navigator's-compartment use. It is of standard panel-mounting size with a 360-degree dial and automatic pointer but includes an automatic 10-minute (1/6th-degree) vernier indicator. It also contains a built-in flag-alarm indicator to warn against trusting the bearing indications when the signals or the equipment operation are unsatisfactory.

The only other item that the pilot need operate is the channel selector switch that tunes the receiver to any one of 100-crystal-controlled channels to select a desired ground station. This switch is on the front panel of the receiver, which is miniaturized to the size of a standard control box and is installed in the cockpit. In this manner, the complexities and dangers of a remote-channel-switching system are avoided. A photographic view of this receiver, approximately 5 by 6 by 7

inches (13 by 15 by 18 centimeters), is shown in Figure 8.

If the standard ID-250 meter is used to display Navaglobe information combined with magnetic-compass information (radio-magnetic-indicator operation), the indicating needle has the property of "pointing to the station" in the manner of an automatic direction finder, specially useful for homing toward the selected ground station. If the standard ID-249 meter is used to display Navaglobe information, the pilot rotates a knob to select a given course, as appearing on the numerical indicator, and uses the vertical bar (deviation indicator) as a left-right meter, as in ominirange-beacon or instrument-approach-localizer practice. The deviation-indicator output could also be used to control an automatic pilot. It is possible that the ID-250 or ID-249 would be used by the pilot, and the special Navaglobe meter with its built-in 1/6-degree vernier would be used for its greater reading accuracy if a radio operator or navigator were aboard.

On observing the Navaglobe bearing indication, the pilot immediately knows his direction from the selected ground station. He can fly to or from the station along this direction by maintaining a corresponding heading, correcting for wind drift so as to keep the Navaglobe bearing indication constant. In the absence of the complementary distance-measuring facility that is intended to be added to the Navaglobe bearing-indication service (see discussion of Navarho in section 5), the pilot may use the Navaglobe bearing indications to fix his position in the same manner as done in radio-compass or short-range omnirange-beacon systems. The outstanding advantages of Navaglobe over airborne direction finders are that Navaglobe has an over-all instrumental accuracy of the order of ± 1/2 degree, and will work reliably through severe noise at ranges approaching several thousands of miles regardless of time of day or season of the year.

The simplicity of the cockpit instrumentation and of the navigational interpretation of the automatic Navaglobe readings makes the system particularly suitable for use in a single- or two-place aircraft, for direct use by the pilot or copilot. This feature has a military significance, for certain types of missions may be carried out by long-range aircraft that carry no special radio operator or navigator.

The remaining portion of the airborne installation consists of the antenna, bearing translator, and power supply. While a simple rod antenna may be used, a special loop antenna has been developed and is shown in Figure 9 without its **streamlined** radome. This is a shielded loop and thus minimizes electrostatic-noise pickup, but it is designed to operate omnidirectionally. The translator and power supply are each housed in a standard A1-D-size assembly. A photograph of the complete airborne Navaglobe installation (less antenna), designated AN/ARN-27 (XA-1), is given in Figure 10. Also being supplied in a special Navaglobe signal generator and simulator for rapid testing and calibrating of the complete airborne installation in the laboratory or in the field. At present writing, the construction of prototype models of the AN/ARN-27 (XA-1) and associated test equipment is completed; after type testing and approval, which is in progress, a moderate quantity of these equipments is to be furnished to the Air Force.

5. Navarho

An ideal system for automatic position-fixing would be one in which combined distance and bearing indications are provided for each ground station, preferably on a single radio-frequency channel. This is the type of service provided by certain very- or ultra-high-frequency medium-range systems operating over land. Distance indications by themselves also offer certain navigational advantages. In recognition of this fact, work is currently being done towards adding distance measurement to the long-range bearing system described in the preceding sections; the complete system has been named Navarho (pronounced Nava·rho).

The presently constructed Navaglobe equipment, ground and airborne, has been designed for the easy addition of automatic distance measurement at such time as airborne frequency sources are sufficiently perfected. The distance measurement involves comparison of the phase of the received Navaglobe signal with a locally produced signal from a highly stable airborne frequency generator. The phase comparison would be performed during reception of the nondirectional synchronizing signal from the ground transmitter.

Such proposals were made a number of years ago by Federal Telecommunication Laboratories, where some research work has since been done along these lines. A high-precision oscillator with counter circuits was developed during 1951, utilizing a commercially available crystal. This work indicated the practicality of distance measurement

through the use of a very-high-precision crystal clock. Certain long-distance flight tests made by the Rome Air Development Center of the United States Air Force in 1953 have further substantiated such expectations.

The problem of adding distance measurement to Navaglobe, resulting in Navarho, is essentially one of manufacturing in quantity relatively small-sized crystal units of the required high stability, which can maintain the requisite stability under the adverse conditions of temperature, shock, vibration, and power-supply variations encountered in aircraft. Oscillators with such stability have been in laboratory use since the second world war. The Rome Air Development Center now has a contractor working on the development of such a crystal oscillator with a specified frequency stability of one part per billion over an eight-hour period; preliminary results presage the successful completion of this development. The expected distance-measurement accuracy is a function of time and for fast aircraft is approximately 1 percent of range.

Plans call for a "building-block" arrangement whereby aircraft may be equipped for bearing measurement, distance measurement, or both. With combined indications, a computer for presenting fix and course information in the most direct terms is practical. As regards the airborne installation, the addition of a special translator unit, constructed along the lines of known techniques, is all that would be necessary to enable the present AN/ARN-27 to provide distance indications. Present procurement plans of the Air Force for Navaglobe ground equipment operating in the preferred and allocated 90-to-110-kilocycle band include arrangements for permitting the future addition of the distance-measuring facility to make a complete Navarho system.

These planned ground stations will also be a higher power than the original 150-kilocycle Navaglobe transmitter at Alamston, which was rated at 10 kilowatts and had relatively small antennas. The experimental 100-kilocycle transmitter at Forestport radiated some 6 kilowatts and with this gave midsummer daytime ranges of 1800 nautical miles (3330 kilometers) and nighttime ranges up to 2600 nautical miles (4820 kilometers) or more. With the planned increase in radiated power, it is expected that sill greater ranges and reliability will be achieved. With even the range demonstrated by the Forestport transmitter, the entire North Atlantic Ocean area could be provided with position-fixing service by means of three or four ground stations. All the oceanic and polar regions of the earth could be similarly serviced by some 30 stations, and since the Navaglobe—Navarho system is characterized by extremely narrow-bandwidth operation, there is sufficient spectrum space for such ground stations within the allocated 90-to-110 kilocycle band.

These points, coupled with the fact that the present Navaglobe equipment is entirely compatible with the complete Navaglobe — Navarho system, indicate that such a system may well fulfill the need for a truly long-range complete radio navigational aid.

The Air Coordinating Committee (ACC) of the United States has recognized the above conclusion through its recently published announcement that the "USAF-developed 'Navarho' System appears to offer the greatest promise" for meeting the operational requirements for long-distance flight operations". The committee proposes that the United States complete the development of this system and perform extensive evaluations both on the Pacific and Atlantic coasts of the United States. It further states that it will invite participation in this program by air and marine carriers, both domestic and international. The committee is currently informing other nations of its position through the International Civil Aviation Organization (ICAO) to promote future standardization of this system should results of its evaluation program of Navarho prove to meet the essential operational requirements. In this connection, the United States Air Force has been conducting preliminary demonstrations of Navaglobe—Navarho equipment for the benefit of the Permanent Council and of the international members of the 5th Session of the Communications Division of the International Civil Aviation Organization.

6. References

1. L. S. Kuter, "Military Air Transport Service", Journal of the Institute of Navigation, volume 2, pages 159-161; June, 1950.

2. R. I. Colin, "Survey of Radio Navigational Aids," Electrical Communication, volume 24, pages 219-261; June, 1947.

3. P. R. Adams and R. I. Colin, "Frequency, Power, and Modulation for a Long-Range Radio Navigational System," Electrical Communication, volume 23, pages 144-158; June, 1946.

4. "Final Acts of the International Telecommunication and Radio Conferences, Atlantic City; 1947:" chapter 3, article 5, paragraph 112.

5. H. Busignies, P. R. Adams, and R. I. Colin, "Aerial Navigation and Traffic Control with Navaglobe, Navar, Navaglide, and Navascreen", Electrical Communication, volume 23, pages 113-143; June, 1946.

6. P. R. Adams and R. I. Colin, "Navaglobe Long-Range Radio Navigational System," Proceedings of the National Electronics Conference, volume 2, pages 288-298; October, 1946.

7. R. B. Murray, "Press Release of the Department of Commerce," Washington, District of Columbia; March 16, 1954.

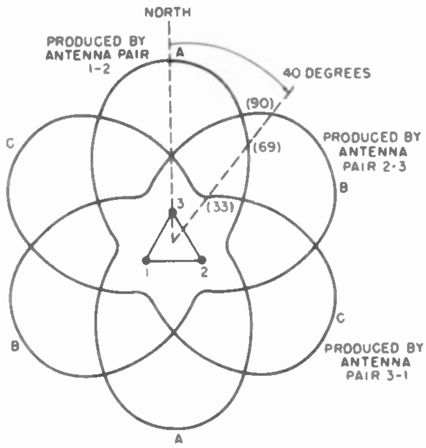


Fig. 1
Navaglobe transmitter radiation patterns. The numbers in parentheses indicate the relative strengths of the A, B, and C signals at the 40-degree bearing. Antennas are at the apexes of the central triangle.

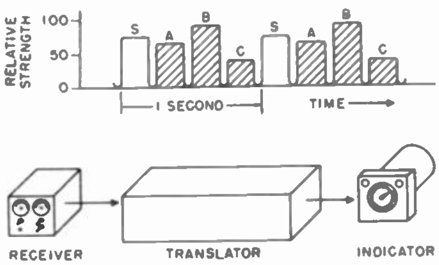


Fig. 2
Operation of Navaglobe receiving system. At the top is shown the detected signals at the 40-degree bearing. The synchronizing signal is at S. In the lower drawing, the receiver detects signals from the selected ground station and applies them to the translator, which in turn identifies them, examines their relative amplitudes, and positions the indicator to the 40-degree bearing.

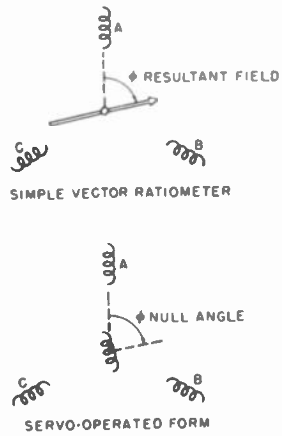


Fig. 3
Principles of Navaglobe bearing translator. In the upper drawing is shown the simple vector ratiometer, in which the field resulting from application of the A, B, and C signals to the respective inductors gives a resultant field direction Ø that determines the position of the magnetic needle. The servo-operated form of indicator shown in the lower drawing uses a moving-coil indicator instead of a magnetic needle. It is equivalent to the upper form but operationally superior.

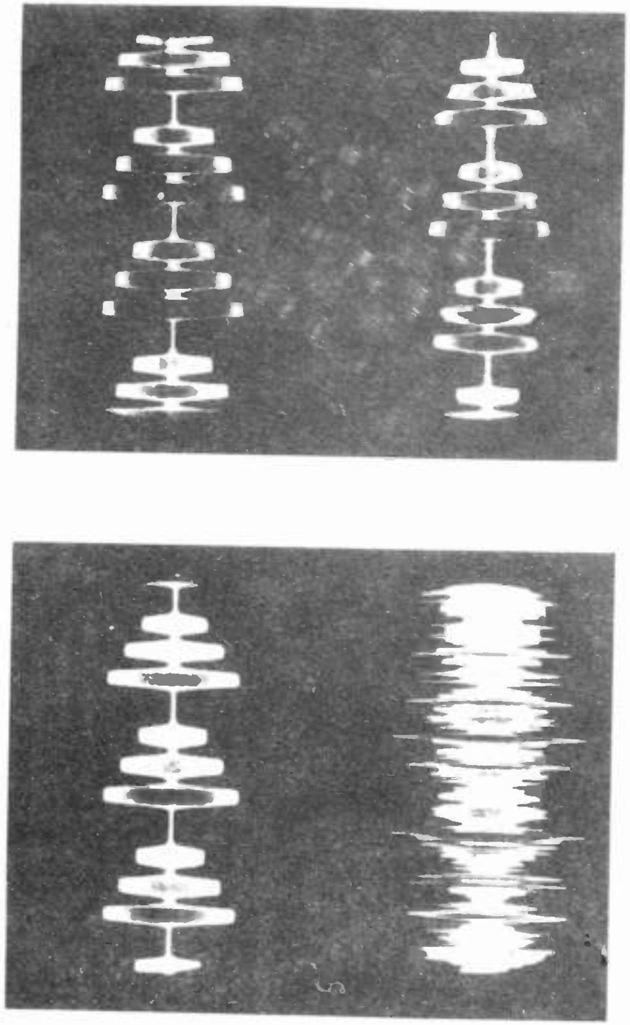


Fig. 4
Oscillograms showing anti-noise behavior of conventional receiver (left) and Navaglobe receiver (right). The upper waveform in each case shows the output when the signal only is applied to the receiver inputs, and the lower waveforms show the outputs when the signal with noise is applied.

Fig. 5 (right)
Experimental Navaglobe equipment used in flight tests.

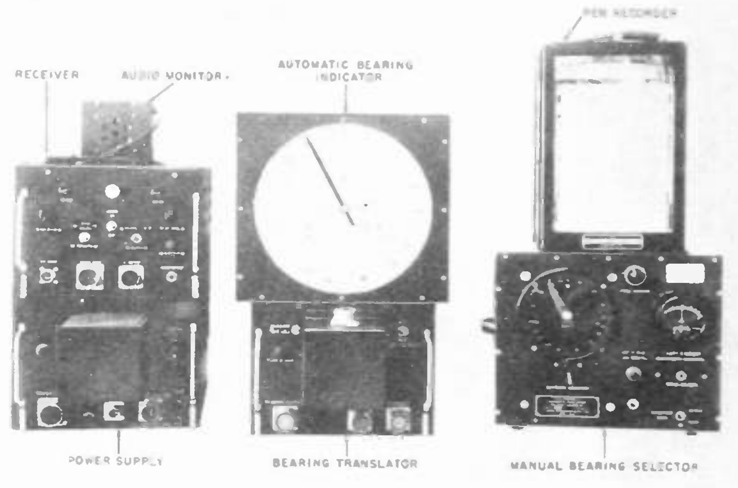




Fig. 6
Map of transcontinental Navaglobe flight-test route. The black triangle marks the location of the transmitter. The dashed line shows the daytime portion of the flights and the solid line the nighttime portion. The maximum range obtained during the day was 1800 nautical miles (3330 kilometers) as shown by point D. The maximum range during the night. shown at point N was 2000 nautical miles (3700 kilometers); this was the extreme limit of the flight.

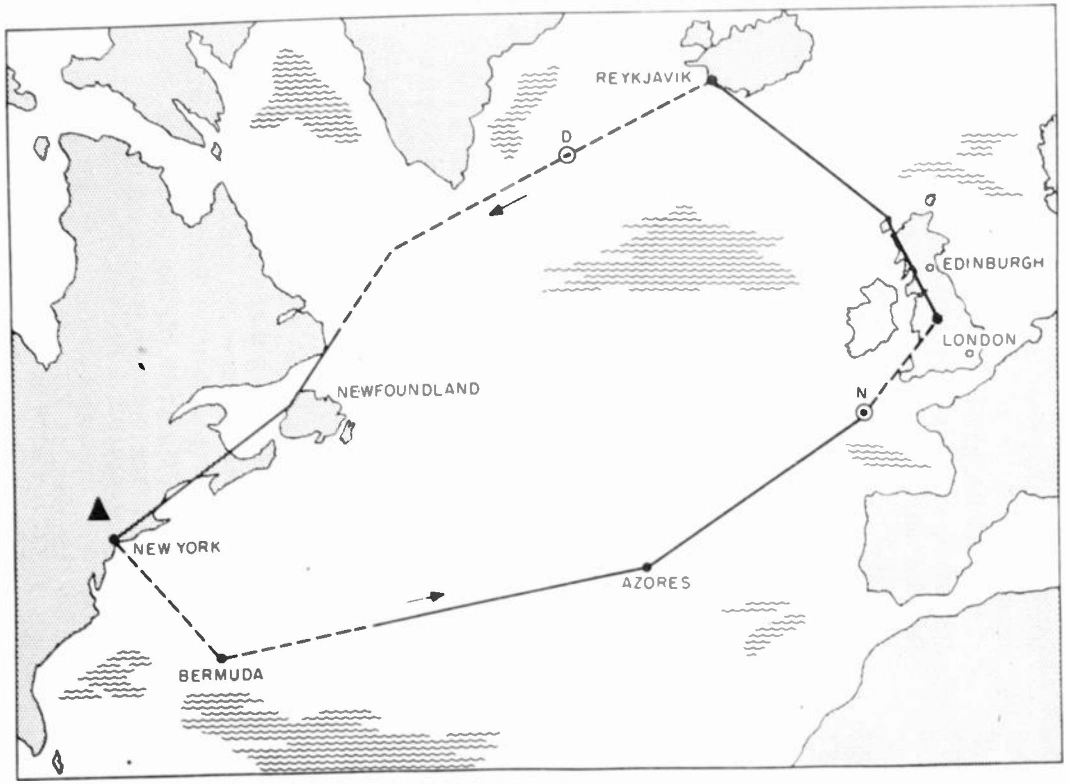


Fig. 7
Map of transatlantic Navaglobe flight-test route. The black triangle marks the transmitter site. Dashed line indicates daytime and solid line nighttime portions of flights. The maximum range obtained during the day was 1800 nautical miles (3330 kilometers), at point D. Point N, 2600 nautical miles (4820 kilometers from the transmitter, shows the maximum range measured during the night, but this was at the onset of daybreak during the flight.

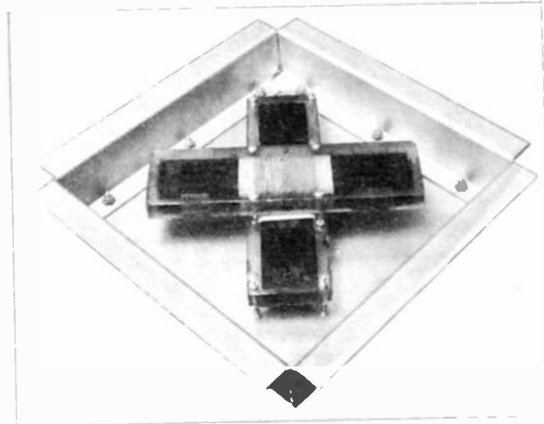


Fig. 9
Flush mounted airborne omnidirectional loop antenna for Navaglobe.

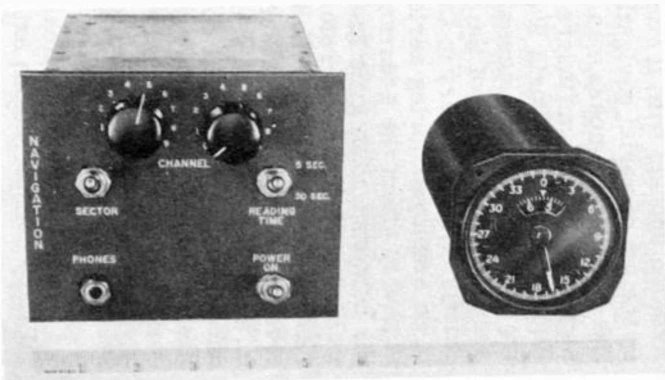


Fig. 8 (left)
The miniature airborne Navaglobe control box and receiver is shown at the left and the vernier bearing indicator is at the right.

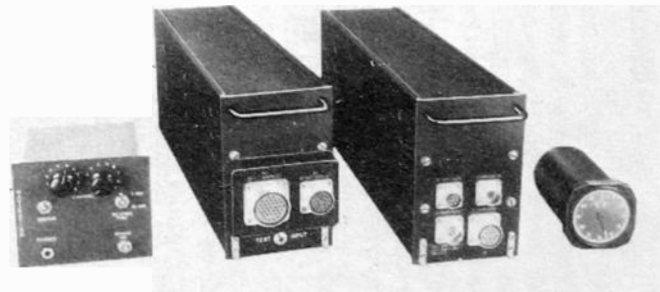


Fig. 10
Photograph of complete airborne Navaglobe installation except antenna. From left to right are the control box and receiver, bearing translator, power supply, and bearing indicator.

Robert C. Rosaler
Kearfott Company, Inc.
Clifton, New Jersey

Summary

Although designed primarily as a compass for polar areas, the N-1 Compass is capable of operation anywhere. The polar navigational problem is described, and the operating principles of the N-1 system, as a solution to the problem, is presented.

Introduction

Charles E. Lindbergh's flight in 1927, over 2500 miles of ocean water, was a marvelous feat. But what is considered even more remarkable to men of aviation is the fact that he navigated over this long distance equipped with only a magnetic compass, augmented by hurried celestial observations. Yet when he reached the coast of Ireland, he was only 3 miles off his charted course!

The art of navigation began when man first sailed his ships beyond the sight of land. Man soon learned that on the open sea his only visible guides were the sun and the stars. He accidentally discovered magnetism and the floating needle, and made a compass. When man began to fly above the clouds, again beyond the sight of land, he could still depend upon the magnetic compass for guidance.

Advances and extension in the scope of aviation brought with it corresponding advances in the art of aerial navigation. Over the years, the magnetic compass was gradually improved and radio direction finders were developed. About the time of World War 1 the principle of the gyroscope was successfully applied. However, the usefulness of the gyro was limited chiefly as an auxiliary to the magnetic compass, to be employed during aerial maneuvering that tended to affect the accuracy and readability of the compass. The gyro itself had a tendency to wander or drift and therefore required frequent resetting to agree with the magnetic compass reading. During World War 11, improvements in magnetic materials and electronic circuit development permitted a continuous electrical link or "slaving" of the gyro to the magnetic compass. But always there was the ultimate dependance upon the earth's magnetic field.

Postwar aeronautical developments and military considerations extended aircraft operations to the Polar Regions. In the polar areas, magnetic compasses are relatively ineffective, due primarily to the weakness of the earth's magnetic field. Radio and radar beacons are few and far between. A dependable, self-contained compass became an urgent need. In 1948, the Kearfott Company, working with the Air Force, developed the answer in the N-1 Compass System.

The Polar Navigation Problem

Before describing the features of the N-1 Compass itself, let us first examine our polar navigational problem from a geographical point of view:

The conventional magnetic compass is always directed toward a particular point on the earth; i.e., the Magnetic North Pole. The location of the Geographic North Pole is determined by adding a known correction to the magnetic heading angle; this correction is known as "variation." The Geographic North Pole has always been the traditional navigational reference. All the meridians converge at the North Pole.

But consider the problem of the polar navigator. Here a convergent meridian system is useless. (Refer to Fig. 1-a). To take an extreme example, one can observe that to fly in a "Southerly" direction from the North Pole can mean an infinite variety of routes. There is no East or West, there is only South, and South in this instance does not define any particular path.

A much more convenient system is the concept of "grid" navigation, which employs a coordinate system or grid as shown in Fig. 1-b. The grid axes are parallel and perpendicular to 000° and 180° meridians of the earth. "Grid North" is defined as the direction going toward the North Pole when at the 000° meridian. One can observe in Fig. 1-b that a constant Grid Heading will define a specific path. It can be proven that flying a constant Grid Heading will always result in a great circle course, the shortest distance between two points on a spherical surface.

A gyro, whose azimuth axis is kept fixed

in space, is fully compatible with the grid concept, for such a gyro automatically provides a true grid reference, provided its azimuth plane is kept tangent to the earth's surface at all times. The angle of the gyro spin axis referenced to this plane will be, at all times, the actual Grid Heading.

However, a gyro whose axis is fixed in space is not rotating with the earth beneath it. To be a useful guide then, the aximuth indication of the gyro must always be corrected for the rotation of the earth, since all useful heading information is referenced to the earth itself. This earth's rotation effect can be easily understood by examining Figs. 2-a and 2-b. In Fig. 2-a the gyro is shown headed Grid North with the 000° meridian shown in its position at 12 noon. However, six hours later, the earth has rotated 90° as shown in Fig. 2-b. The gyro now has an apparent Grid Heading of 90°. This effect is usually referred to as "apparent drift", but could better be described as "earth's rotation effect." The apparent grid reference or Grid Heading of a gyro located on the equator will not be affected by earth's rotation. "Earth's rotation effect" is a function of the sine of the latitude where the gyro is located, and as such is a calculable value.

The principles of operation of the N-1 Compass now can be concisely stated in the light of the basic navigational problem that we have just examined:

First, a nearly drift-free gyro to provide a dependable space reference.

Second, a computing device and indicator to calculate and correct for the effects of earth's rotation and to display the resultant heading on a suitable dial.

Third, a servo system to keep the gyro azimuth plane tangent to the surfact of the earth.

Fourth, appropriate electronic circuitry to permit the system to operate either as a conventional magnetic-field guided system or as an independent directional reference, and to provide power distribution and appropriate control functions.

Description of the System

Figure 3 illustrates the components that make up the N-1 Compass System. A brief description of each of these components follows:

The Directional Gyro

The Directional Gyro is the primary directional reference when the system is in Directional Gyro operation. When the system is in Magnetic-Slaved operation it acts as a stabilizing directional element. The gyro-motor is mounted within an inner gimbal which in turn is pivoted within an outer gimbal. Figure 4 illustrates the gyro arrangement in elementary form. The axis of the gyro-motor, i.e., the gyro spin-axis, points to the reference direction. A synchro type transducer monitors the position of the Inner or Leveling Gimbal. Another synchro measures the angular position of the outer or Azimuth Gimbal. This synchro delivers to the rest of the system an electrical signal proportional to the azimuth position of the gyro. The gyro spin axis is always kept level in the "case horizontal plane" as follows: Assume that the spin axis is momentarily off level. An error voltage is induced in the leveling synchro which, through an amplifier, powers a leveling torquer motor on the azimuth axis. The application of torque on the vertical axis will precess the inner leveling axis until it is case level and the leveling synchro signal is reduced to a null. In this manner, the gyro spin axis is always kept, on an average, in a plane tangent to the surface of the earth. An azimuth dial attached to the vertical axis gimbal is visible through a window in the upper part of the gyro housing. This is to enable observation of the azimuth drift rate during system tests and maintenance.

The entire gyro is hermetically sealed and is filled with nitrogen. In addition, the directional gyro motor itself is hermetically sealed in a separate container. Random drift of the gyro is minimized by performing final balancing of the gyro from the outside, through the hermetic seal, by means of a Hermeflex*, a unique rotary-bellows device. The gyro case is supported in a mounting base by three shock mounts.

Type C-2 Remote Compass Transmitter

This unit is the magnetic-direction sensing component of the compass system when operating as a magnetic slave compass. It is, in a sense, a magnetic compass in itself, transforming the magnetic field vector into electrical signals proportional to the direction of the magnetic field. The magnetic sensing element is pendulously mounted so that its average position in the

---

*Copyright, Kearfott Company, Inc.

horizontal component of the earth's magnetic field is normally maintained, regardless of the attitude of the aircraft. The hemispherical ball is filled with fluid to damp any swinging of the element.

Amplifier

The amplifier is the "distribution center" of the system. All the other components in the system are connected to it. Primary power to operate the compass is fed to the amplifier and distributed to the system's components. In addition, provision is made for obtaining remote indications of azimuth data to any other equipments that may require it. Azimuth, Correction and Leveling signals originating in the various components of the system are each received, amplified and transmitted by separate channels in the amplifier. Economy of space is accomplished in the design by utilizing many channels for dual purposes. Three channels of the amplifier and the power supply in the amplifier are assembled in two chassis, each of which is attached by one of its sides to a distribution panel board. One of the chassis is hinged so that it can be swung away from the other to provide access to the chassis components. The two chassis and the panel board are housed in a cabinet which is shock mounted by a cradle type mounting frame. The physical construction is extremely rugged and is designed for minimum maintenance under the severest military conditions. Conservative use of tube ratings have resulted in an outstanding record of maintenance-free operation.

Master Indicator

The Master Indicator is "the flight instrument" of the system. It coordinates the data received from the gyro and the computed corrections, and displays it in useful form to the navigator. The Master Indicator contains the correction computer for correcting earth's rotation effect, and also generates the remote synchro signals for operating other equipments, such as remote indicators, radar or automatic pilot. Fig. 5 is a close-up view of the indicator as seen by the navigator. The Latitude Correction Control Knob, in the upper righthand corner of the indicator, provides a means for selecting either Magnetic-Slaved operation or Directional Gyro operation of the compass system, and also sets the proper latitude correction rate. The latitude dial is mechanically linked to this knob. When the knob is rotated counter-clockwise until the pointer is in the "off" position, the system is then in Magnetic-Slaved operation. For Directional Gyro operation, the pointer is set to the local latitude of the aircraft. The setting of this latitude dial controls the correction rate to be made to the Directional signal. The Synchronizer Control Knob, at the lower righthand corner of the indicator, provides a means for quickly synchronizing the indicator heading pointer with the correct magnetic heading, when the system is in magnetic slaved operation. It also provides a means of setting the heading pointer to the desired heading reference when the system is in Directional Gyro operation. An annunciator pointer indicates the proper direction to rotate the indicating heading pointer in order to synchronize. The Correction Servo Indicator is a rotating white dot, which tells the observer that the proper correction (either latitude correction or magnetic heading correction) is being applied to the system. An adjustable cam compensator is provided in the back of the indicator to compensate for transmission error deviations at 15° points. The entire indicator is hermetically sealed and filled with nitrogen. Rotation of the knob shafts through the hermetic seal is again facilitated by a Hermeflex.*

Slaving Control

The slaving control is a single degree-of-freedom gyro whose axis is positioned in proportion to the turning rate of the airplane. During severe turns of the aircraft, the magnetic compass transmitter may be subject to errors if the turn is of long duration or at a high rate. In addition, because of case leveling of the gyro, long-duration turns will cause leveling to the dynamic vertical as determined by the aircraft making the coordinated turn. The Slaving Control electrically cuts off the leveling action in the Directional Gyro when the system is in Magnetic Slaved or Directional Gyro Operation, and also opens the signal circuit from the magnetic direction sensing component (the C-2 transmitter) when in Magnetic Slaved Operation. As soon as the aircraft returns to straight and level flight, the slaving control closes these circuits, returning the system to normal operation. This unit is also hermetically sealed and filled with nitrogen.

Type V-7A Gyro Magnetic Compass Indicator

This indicator repeats the compass reading in a remote protion of the aircraft as required. A manually operated setting knob is provided at the front of the indicator for rotating the dial 360° in either direction without changing the indication of the pointer itself. A large dial is employed to assure maximum readability.

This unit is hermetically sealed and filled with nitrogen.

Type V-8 Gyro Magnetic Compass Indicator

This is functionally identical to the V-7A indicator except that the dial is smaller. This indicator would be employed where less space is available and a less accurate compass reading is sufficient.

Principles of Operation

The basic azimuth circuit is shown in Fig. 6 in block form. An electrical signal corresponding to the azimuth position of the gyro is transmitted by azimuth takeoff (Synchro) B-205 to azimuth control transformer B-302. An error signal from B-302 positions the Master Indicator Heading Pointer via the azimuth channel of the amplifier and azimuth servo motor B-306. The pointer rotation continues until the pointer indicator matches that of the signal of azimuth take-off B-205. When the Synchronizer Control Knob is rotated, a mechanical linkage between the control knob and the stator of azimuth control transformer B-302 rotates the stator. This displaces the output signal of the control transformer from null and causes the Master Indicator Heading Pointer to rotate until a null is restored in the control transformer. The angular rotation of the pointer will be proportional to the angular rotation of the synchronizer control knob.

Directional Gyro operation is shown in Fig. 7. Correction for earth's rate is introduced into the azimuth circuit by rotating the stator of Azimuth Control Transformer B-302. The Latitude Correction Control Knob is rotated until the Latitude Correction Pointer indicates the desired latitude. This simultaneously positions Latitude Correction Transmitter B-307. The magnitude of the voltage output of B-307 determines the speed of the correction servo motor signal generator MG-301 via the correction channel of the amplifier and the associated feedback circuitry. MG-301 rotates Azimuth Control Transformer B-302 at a rate corresponding to the latitude set in.

Figure 8 outlines Magnetic Slaved Operation. The Directional Gyro signal is continually being monitored by the C-2 Magnetic Compass Transmitter. Any change in gyro heading due to drift would momentarily cause the Master Indicator Heading Pointer to change correspondingly via the azimuth control circuit. However, this would also rotate Magnetic Detector Control Transformer B-301 (mechanically linked to pointer) which in turn is being energized by the unchanging signal from the Remote Compass Transmitter. This error voltage will rotate Azimuth Control Transformer B-302, via the correction channel of the amplifier and signal generator MG-301, until the error in B-301 is nulled. At that time the Master Indicator Heading Pointer will return to its position corresponding to the heading as indicated by the C-2 Compass Transmitter. During turns of the aircraft, the Master Indicator Heading Pointer is servoed rapidly to its new position by the azimuth circuit.

The correction rate of the magnetic correction circuit is very slow compared to that of the basic azimuth correction circuit. During the passage of the aircraft over areas of magnetic distortion, this slow correction rate would limit any magnetic error creeping into the system. A slow rate also acts to integrate or smooth out the effects of oscillation of the magnetic element. The relatively fast response of the azimuth channel will keep the Master Indicator pointer at substantially the same position. As soon as the area of magnetic distortion or any cause of magnetic error has been cleared, the servo will restore the Master Indicator reading to that of the magnetic transmitter. This feature also permits the proper heading to be displayed during turns of the aircraft, even when such turns affect the signal output of the Remote Compass Transmitter. During very rapid turns of the aircraft, (exceeding 3° /min.), the Slaving Control cuts out the C-2 Compass Transmitter and the leveling circuit of the gyro, thereby reducing the chances for erroneous indications.

Conclusion

Today many N-1 Compass Systems are successfully flying in aircraft all over the world. The drift rate of the system never exceeds 1 $1/2^o$ per hour. Since this drift is of a random nature, the average system error over long flights is much less, being on the order of $1/2^o$.
Air Force Data show a remarkable history of dependable, trouble-free operation. The high performance and low-maintenance record of this system can be attributed to three principal factors:

1. The use of special electro-magnetic components such as synchros, servo motors and tachometer-generators. The characteristics of the components required for this system were such that they were not available commercially at the time the system was designed. Intensive design-development led to a line of electro-

magnetic components that were sufficiently accurate and dependable for the requirements of the N-1 System; as, for example, a synchro having a maximum error of 5' of arc, yet having an overall diameter of only 1 1/16". These required the solution of many problems relating to the design and quantity production of these high performance components.

2. The extensive use of hermetic sealing of all mechanical assemblies. Only a true metal-to-metal hermetic seal can be depended upon to keep foreign matter permanently out of the mechanisms. Oxidizing of the lubricants, as well as all forms of corrosion, are completely eliminated by filling the hermetically-sealed containers with an inert dry gas such as helium or nitrogen.

3. All azimuth corrections are performed upon the azimuth signal rather than upon the gyro itself. This feature keeps the gyro simple in construction, free of extra accessories that might otherwise compromise the gyro's basic function. Power input to the gyro is thereby kept at a minimum.

The N-1 Compass System is a giant step in the progress of navigation instrumentation. It is the fore-runner of smaller and lighter directional gyroscope systems of comparable performance that are now being developed by the Kearfott Company for lightweight military and for commercial aircraft.

The Kearfott Company wishes to express its appreciation to the Wright Air Development Center of the U. S. Air Force for permission to present this paper.

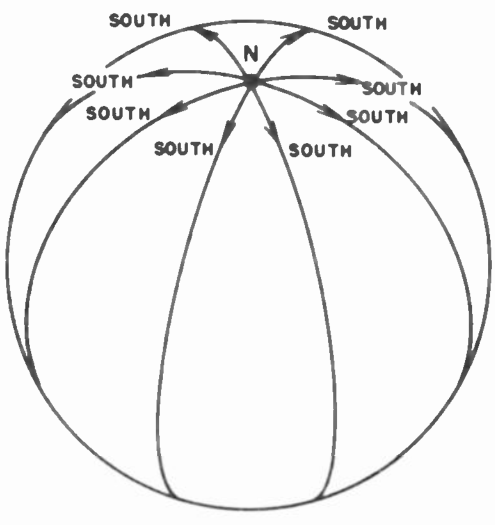


Fig. 1(a) - Everywhere you go, it's South.

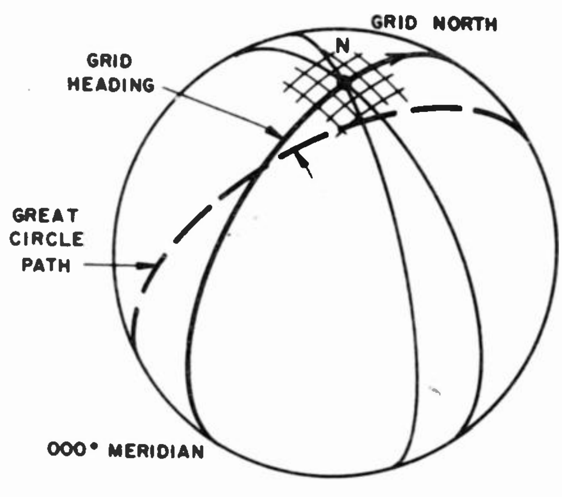


Fig. 1(b) - The grid heading concept.

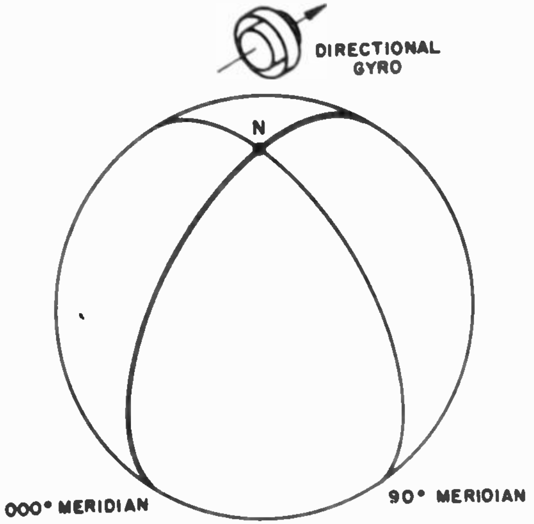


Fig. 2(a) - At 12 noon, grid heading is 000°.

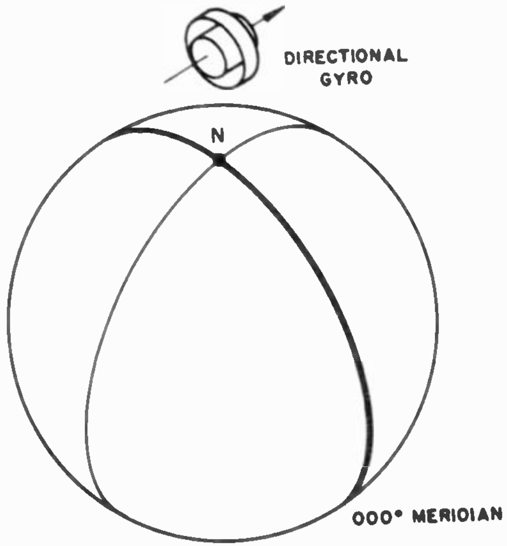


Fig. 2(b) - At 6 p.m., grid heading is 90°.



Fig. 3
(1) directional gyro; (2) USAF Type C-2 remote compass transmitter; (3) Amplifier; (4) master indicator; (5) slaving control; (6) USAF Type V-7A indicator; (7) USAF Type V-8 indicator.

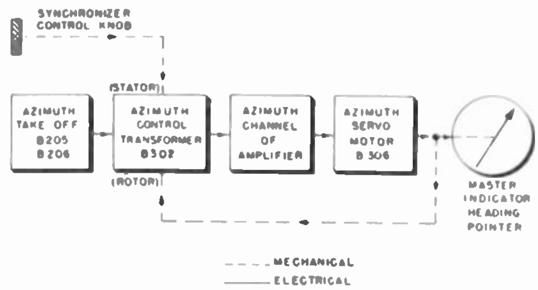


Fig. 6 - Azimuth circuit - block diagram.

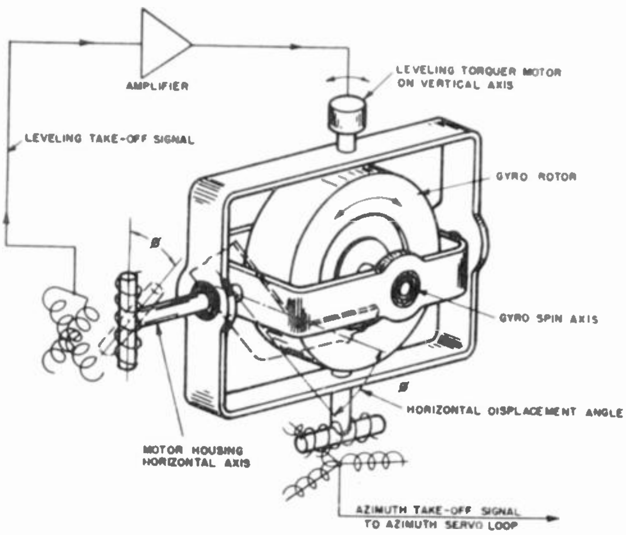


Fig. 4 - Operation of leveling servo loop.

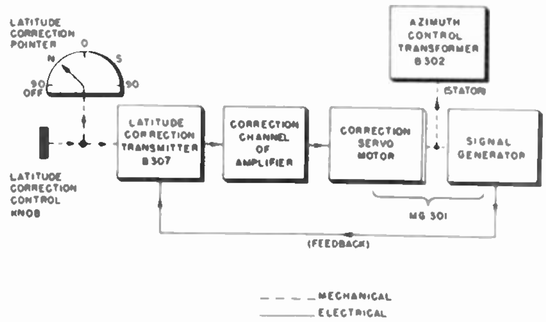


Fig. 7 - Directional-gyro operation - block diagram.

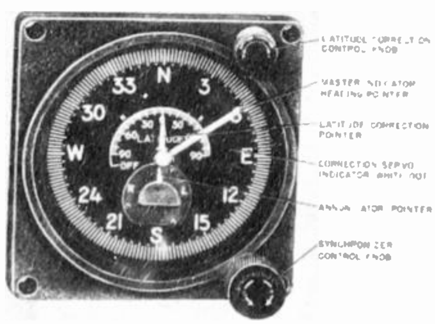


Fig. 5 - Master indicator - front view.

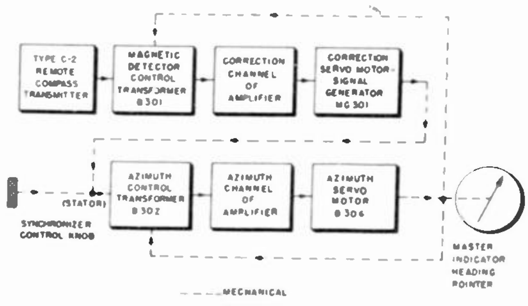


Fig. 8 - Magnetic-slaved operation - block diagram.

A 227 MC PULSE POSITION MODULATION TELEMETERING UNIT

D. G. Mazur
Naval Research Laboratory
Washington 25, D. C.

Summary

Operation and performance of a 15 channel PPM telemetering unit is described. Designed for use in medium size rockets and to be compatible with existing ground equipment, the present transmitter has a power output of 10 watts peak during the 3μs pulse. Sampling rate is nominally 312.5 cps but one channel may be gated four times during each frame, yielding a sampling rate of 1250 cps. Overall accuracy has proved to be within one per cent and performance has been excellent in 10 rocket flights. Commercial units will have increased power and crystal controlled rf.

- - - - - - - -

Since 1946, the Naval Research Laboratory has been engaged in a program of upper atmosphere research using V-2, Viking, and Aerobee rockets launched at the White Sands Proving Ground, Las Cruces, New Mexico. A vital part of this program has been the continued development of pulse position modulation telemetering systems which would meet the exacting requirements for relaying basic research measurements. These are accuracy, reliability and high percentage recovery of noise-free data.

This paper describes the 15 channel AN/DKT-7 ( ) telemetering transmitter (Fig. 1) which was primarily designed for use in the Aerobee rocket and is compatible with existing NRL ground recording stations, the AN/FKR-1( ) equipment.

The specifications of the transmitter are as follows:

15 channels each with a normal sampling rate of 312.5 cps.

Overall intelligence rate 4700 samples per second.

Input range 0 to 5 volts positive.

RF pulse width 3 microseconds.

RF frequency 227 mc.

Peak power output during the pulse 10 watts.

Channel deflection range 150 microseconds.

Accuracy within 1% relying on in-flight calibration.

The transmitter has 15 input or data tubes, each data tube containing a triode-duo diode in its envelope. The triode section essentially acts as a cathode follower with the data to be transmitted connected to the grid. When either diode plate is held sufficiently positive however, the current flow to the diode plate raises the cathode of the tube to a point where plate current cutoff results (Fig. 2). Appropriate gating voltages on the diode plates thus permit the triode sections to be turned on and off in sequence. The sequence is achieved by using combinations of a scale of 16 counter outputs as gating voltages. The resultant data current for a given input is collected at a common point and transformed into a voltage which is compared to a sawtooth to provide pulse-width modulation.

The gating is done in two steps. First, every fourth data tube has its plate tied together, that is, channels 1,5,9,13 have a common output. In similar fashion channels 2,6,10, and 14 have a common connection. There are four such groups, the last one having only three data tubes. Simultaneously, the first channel tube of each group is allowed to conduct for 800 microseconds. This is achieved by having the diode plates of these tubes at ground potential during this period. Then the second channel tube of each group is allowed to conduct for 800 microseconds, and so on.

Each of the four groups is connected to its own collector tube which is, in turn, gated in similar fashion. However, the collector tube gating is such that the first tube of the first group conducts, then the first tube of the second group conducts, and so on. Fig. 3 shows a mechanical analogy of the commutation process. By means of this cascade arrangement, sequential turn-on of the data tube current is obtained; and, by tying the collector plates to a common resistor, the current derived from the voltage applied to each channel may, in sequence, influence the operation of the pulse-width modulator.

Fig. 4 displays the interconnections of the data and collector tubes and the gating combinations which are applied to the diode plates. The potential at the plate of the data tube is set by the grid potential of the collector tube, which is fixed. This permits the tube to act as a cathode follower during the sampling time.

The gating is done using the complementary outputs (Fig. 5) of a scale of 16 triode counter in various combinations. Each data tube has its diodes connected to the scale of 8 and 16 counter outputs while each collector tube has its diodes connected to the scale of 2 and 4 counter outputs.

It should be noted that the maximum time alloted for each channel is 200 microseconds. This spacing is derived from a free running 5 Kc Clapp oscillator. The oscillator drives a thyratron repetitive sawtooth generator arranged to both generate a sawtooth and simultaneously a counter drive pulse. The counter triggering pulse occurs in coincidence with the discharge of the sawtooth, giving synchronism between the switching of the data tube current and the use of the sawtooth.

Pulse-width modulation is accomplished in a regenerative, diode-coupled, voltage comparator circuit (Fig. 6). The sawtooth is coupled into a pentode and causes its cathode to rise. Meanwhile the triode has been conducting and its plate is low. The sawtooth eventually reaches an amplitude sufficient, when coupled through the diode, to raise the cathode of the triode and initiate cut-off. This cutoff is regeneratively aided by the plate voltage drop of the pentode coupled back to the triode. The triode remains cut off until the sawtooth discharge, when initial conditions reassert themselves. The data current is introduced into the plate grid coupling network and has the effect of modifying the voltage at which the triode will be cut off. Large data currents will lower the grid potential of the triode, and result in turn-off at lower values of sawtooth voltage. Conversely, small data currents will raise the triode's grid potential, and result in turn off at higher values of sawtooth voltage. In practice, the sawtooth voltage and an adjustable resistor in the triode grid circuit are varied so that there is a finitely wide triode output-pulse for each channel tube at zero input.

Of the 200 microseconds alloted for channel spacing, only 150 microseconds are used for full voltage modulation. Part of the remainder is then left for the finite channel pulse widths, or guard bands, and part for the deionization time of the sawtooth generator. In addition some of the period has to be wasted due to poor rise time of the switching currents.

A data tube may have its sampling rate quadrupled by removing the connections from the scale of 8 and 16 counters, normally tied to its diodes. In this fashion, the data tube conducts four times during each frame instead of once. Of course, the remaining three data tubes which are normally connected to a common point must be removed, and the single high speed channel conducts during its own as well as their time period. If channel three is made a high speed channel, conduction will occur during periods 3, 7, 11 and 15. Thus at a sacrifice of three channels, a high speed channel may be created with a sampling rate of 1250 cycles. It should be noted that this is not the same as connecting four normal channels to the same input; since, due to variations in channel guard bands, four normal channels would have different zero positions, while the high speed channel has the same zero each time it conducts. Removal of the 8 and 16 counts is easily effected by clipping the diode pins of the data tube. Recording of the high speed channel must be done in special fashion using an oscilloscope with a repetitive synchronized trigger in order to gain the advantage of the increased sampling rate.

The pulse-width modulation output of the voltage comparator circuit is differentiated and mixed with frame synchronization pulses in a common stage.

The frame synchronization, necessary for locking the ground decoding and recording equipment to the airborne pulse pattern, is obtained from a triple pulse code consisting of three pulses spaced approximately 7.9 microseconds apart. The characteristics of this code was dictated by the design of the ground equipment, which had already been in existence when the AN/DKT-7( ) unit was developed. The generation of the triple pulse in the transmitter is adjusted to occur 100 microseconds following a channel reset. The delay prevents the possibility of any data pulses combining together so as to give a false synchronization, and gives the ground station a clear period where nothing but the synchronizing signal is being transmitted.

The mechanism of synchronizing code generation is initiated by the scale of 16 positive counter output. Prior to this time all of the counter tubes are in the reset position, and this condition occurs once every 16 driving pulses. The scale of 16 positive output triggers a phantastron delay generator which is adjustable to 100 microseconds. Its output pulse is then connected to trigger a self-restoring multivibrator having in the plate of the normally conducting tube an LC resonant circuit. Upon being triggered, this conducting tube shuts off and shock excites the LC circuit into oscillation, producing a damped sine wave whose peaks are separated by 7.9 microseconds. Constants of the multivibrator may be varied to produce two, three, or four cycles of the sine wave before restoring itself. The sine wave is then shaped into pulses and mixed with the differentiated pulse-width modulation output in a video amplifier stage.

Since the whole channel period following counter reset is used for the triple pulse, only 15 intervals remain for data pulses. The second interval is called channel 1, the third channel two, and so on.

The combined video is fed to a blocking oscillator having a three microsecond delay line in its circuitry to shape the pulses. A self excited Colpitts oscillator operating at 227 mc is grid pulsed by the blocking oscillator output. A simplified version of the rf oscillator is shown in Figure 7. Driving pulses are applied to the grid and overcome the self bias developed by previous pulses. The time constant is adjusted to maintain the oscillator in the off position between pulses, and in the absence of modulation the oscillator will squedge at a low rate. The output RF pulse is nearly square. Peak power during the pulse is 10 watts with an average power of 0.2 watts.

The accuracy of the AN/DKT-7( ) unit is enhanced and maintained by use of in-flight calibration. The transmitter provides a source of constant current to the calibration circuitry (Figure 8). This source consists of a regulated cathode follower having its grid potential, as well as its plate potential, set by voltage reference tubes. An adjustable resistor, located in the calibrator and connected in series with the cathode of the cathode follower, permits 10 milliamperes of current to flow through five precision 100 ohm resistors, resulting in one volt tap points from zero to five volts.

The calibrator itself (Figure 9) consists of a motor-operated set of cams controlling micro-switches which in sequence, disconnect the data from each channel. The channel is then connected to a commutator which samples the six precision divider tap points. A complete calibration period for each channel lasts for one-third second and during this period every channel has applied to it zero, 1, 2, 3, 4, and 5 volts briefly. All 15 channels are calibrated in sequence and it is possible to delete calibration of any channel if so desired. The calibration cycle repeats itself every 16 seconds and about two percent of the data is lost during calibration time. The circuitry is stable to much better than one percent over widely varying conditions of the input voltages.

Primary power sources for the unit are obtained from an eight battery which furnishes filament voltage, and a twenty-eight volt battery which runs both a dynamotor and the calibrator motor. The dynamotor supplies a nominal 320 volts DC which is dropped in successive stages to provide suitable plate voltages for the various circuits. Susceptible circuits are operated from a regulated 150 volt bus.

The size of the transmitter is approximately 9" high, 9" wide, and 12" deep. It is housed in a pressure tight case, all panel holes being sealed with gaskets or O-ring seals. AN connectors for power, control, and input leads are mounted at right angle to the front panel to conserve installation space. The transmitter weighs 18 pounds, the external calibrator 3½ pounds, and when powered with lead acid batteries, the complete installation, less antenna and interconnecting cables, weighs about 43 pounds. Use of Silvercels instead of lead acid batteries reduces the overall weight of the equipment to about 32 pounds.

The unit is fabricated of aluminum and consists of two decks hinged together in back (Figure 10). Interdeck wiring is connected through a plug and receptacle arrangement with an extension cable being used for servicing in the open position. The premodulator deck, containing the commutation and timing circuitry, is fastened to the bottom of the front panel; while the power deck is hinged to it and has mounted on it the blocking oscillator, the RF oscillator, and power circuits. The front of the power deck is secured to the front panel by a captive screw. Miniature turrets, attached to the tube sockets, support most of the electrical components of the premodulator deck. This type of construction, although a bar to quick replacement of all components, permits maximum utilization of the space. All tubes are of the miniature type with exception of the RF oscillator where a subminiature is used. The decision to standardize on miniature tubes was based on the necessity for using a miniature triode duo-diode, the ease of replacement of most tubes, and the emphasis placed on reliability.

Heat conduction in the premodulator portion is aided by the use of heavy T-plate type construction shown in Figure 11. This helps eliminate any source of hot spots and is conducive to an even temperature throughout. Heat is conducted through the T-plates to the front panel and thence to the mounting structure. No shock mounts are used in or for the equipment, although the premodulator deck tubes are retained by aluminum plates which have sponge rubber spacers between them and the tube tops.

Detailed performance analysis was made of this equipment subsequent to its design and construction. The following electrical characteristics are some of those checked:

1. 5 kc oscillator drift $\pm$ 1 cps with 10 volt plate variation, $\pm$ 5 cps drift with heat, operating 15 minutes in the case.

2. Modulation linearity within 1% of a straight line up to four volts and within 2% from four to five volts.

3. Maximum crosstalk about 0.5 microseconds or 0.3% of the total deflection.

4. Channel noise or jitter is less than 0.3 microseconds or 0.2% of the total deflection.

5. Calibration accuracy is within 0.2% using 0.1% calibrator divider resistors.

6. Variation of pulse position with varying input impedance is in the order of 0.3%.

7. Normal channel time constant is 500 microseconds.

8. Repeatability, obtained by plotting successive calibration points throughout several rocket flights, is within ½%.

9. Pulse position drift with heat over a 15 minute period is held to within 1%.

10. Overall accuracy is about 2% without flight calibration and better than 1% with flight calibration.

11. Overvoltage protection is not required for negative voltages, and only required for positive voltages in excess of five or six times the normal input voltage range.

12. Battery operation time is 44 minutes on lead acid batteries and 38 minutes when powered with Silvercels.

The AN/DKT-7 ( ) transmitter has been flown in ten Aerobee rockets. A typical nose installation is shown in Figure 12. Peak altitudes of these rockets have ranged from 50 to 86 miles dependent on the instrumentation payloads. Maximum accelerations encountered are approximately 15 G. There have been no failures to date. Recovery of completely noise-free data has been better than 96% of the total flight time on each flight, with most of the losses occuring as the rocket rose through the launching tower, as a result of antenna shadowing. In all instances every transmitter prepared as the flight transmitter has been flown, and in no case has a spare transmitter been resorted to at the last moment.

This equipment is now in commercial production by General Electronics Laboratories, Inc., Boston, Massachusetts with the nomenclature AN/DKT-7(XN-1). The commercial unit is shown in Figure 13. The XN-1 version is approximately the same size and weight, as the NRL transmitter and utilizes the same type of construction. The premodulator deck is identical; however, the rf and blocking oscillator circuits have been redesigned to afford crystal control operation, higher peak power output, and minimum spectrum radiation. The XN-1 units employ a 5687 trigger amplifier and blocking oscillator, and a 5687 modulator amplifier A 5703 is used as the overtone crystal oscillator at 75.833 mc. and is tripled to 227.5 mc using another 5703. A third 5703 tube constitutes a driver stage feeding two 6021 tube push pull rf amplifiers. The tripler, driver, and final amplifiers are plate pulsed by the 5687 modulator amplifier. In all other respects the normal tube complement of the AN/DKT-7( ) is used.

Peak power output of the XN-1 unit is better than 40 watts and pulse shaping has b en employed to reduce radiation at points 1.5 mc and beyond from the carrier to 30 db. below that of the carrier. These units are now coming off the production line and are to be used operationally starting in May.

Two versions of the AN/DKT-7 transmitter have been described. The NRL units have had excellent performance in ten Aerobee flights, proving to be reliable, accurate and rugged. They have transmitted data to better than one percent accuracy with good noise-free continuity in rocket flights reaching up to 86 miles in altitude. The commercial units are designed to retain all of the desirable characteristics of the original transmitter and, in addition, satisfy rocket range requirements for frequency stability and minimum carrier spectrum used. The higher output power should result in the same safety factor of signal to noise up to 190 miles.

The development and production of the AN/Dkt-7( ) transmitter was carried on under the direction of J. T. Mengel. Those responsible for the design, development, construction, and field use of this unit were K. M. Uglow, N. R. Best, R. Lowell, R. Freudberg, J. B. Flaherty, J. Y. Yuen, L. F. Schmadebeck, and the author. Operational assistance has been provided in all rocket flights by members of PSL, New Mexico College of Agriculture and Mechanic Arts under contract to the Naval Research Laboratory.

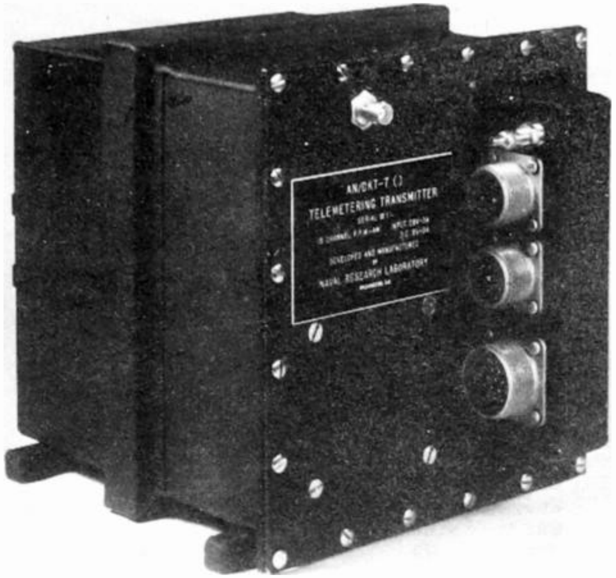


Fig. 1 - AN/DKT-7( ) transmitter.

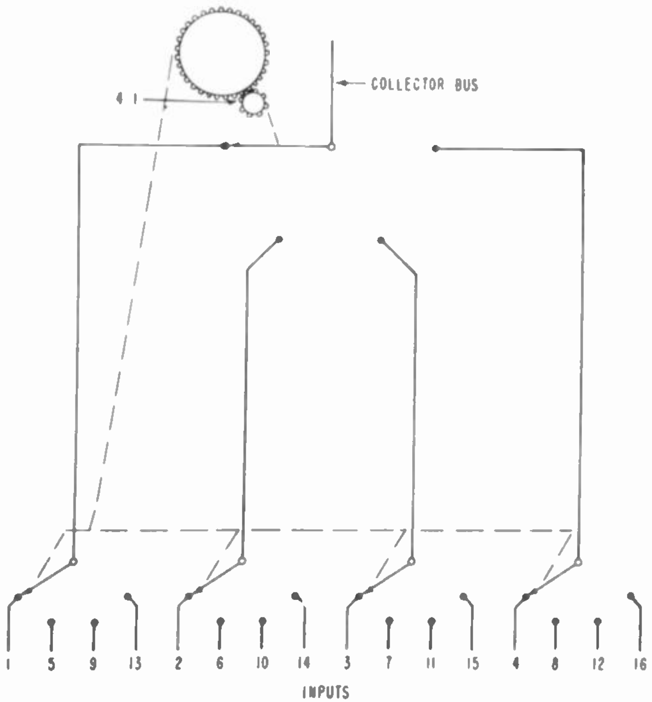


Fig. 3 - Mechanical analogy of commutation process.

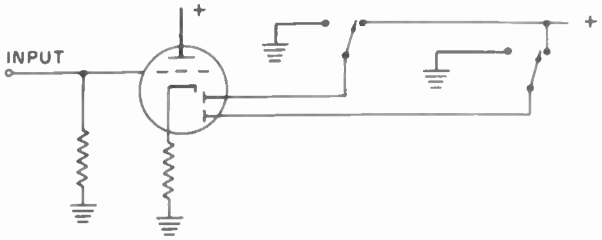


Fig. 2 - Basic data tube circuitry.

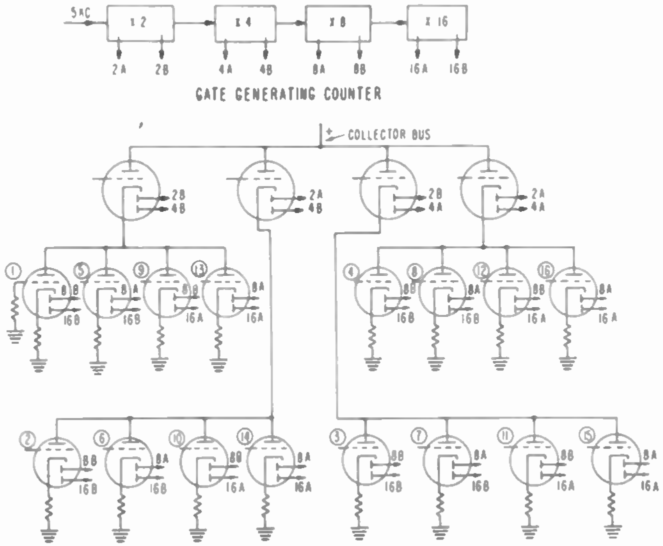
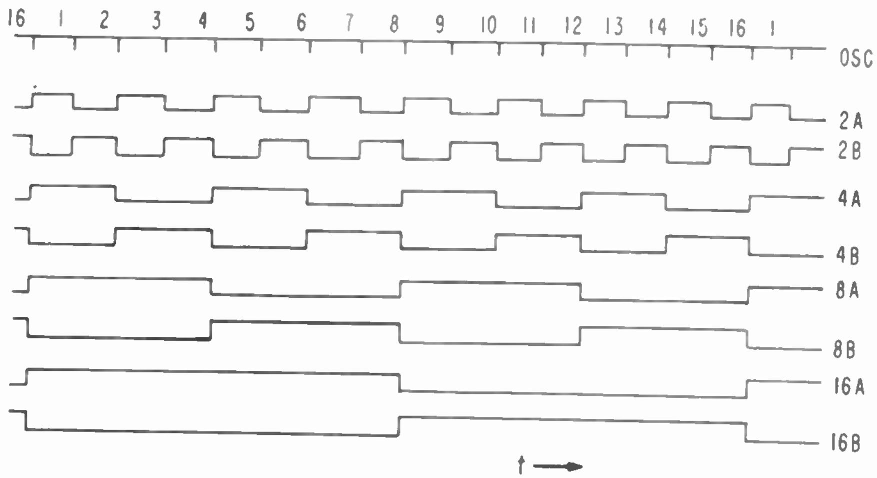


Fig. 4 - Electronic commutation connections.



Counter Plate Waveforms

Interval	Diodes Connected for Coincidences			
1	2B	4B	8B	16B
2	2A	4B	8B	16B
3	2B	4A	8B	16B
4	2A	4A	8B	16B
5	2B	4B	8A	16B
6	2A	4B	8A	16B
7	2B	4A	8A	16B
8	2A	4A	8A	16B
9	2B	4B	8B	16A
10	2A	4B	8B	16A
11	2B	4A	8B	16A
12	2A	4A	8B	16A
13	2B	4B	8A	16A
14	2A	4B	8A	16A
15	2B	4A	8A	16A
16	2A	4A	8A	16A
1	2B	4B	8B	16B

Fig. 5 - Counter plate wave forms.

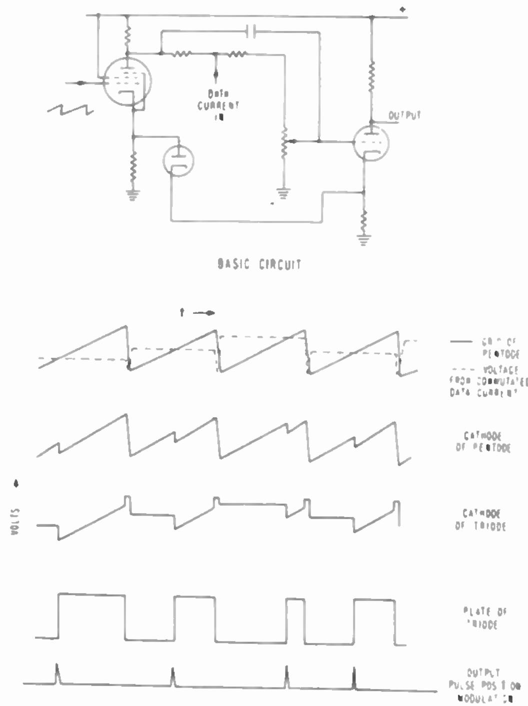


Fig. 6 - Voltage comparator circuit and wave forms.

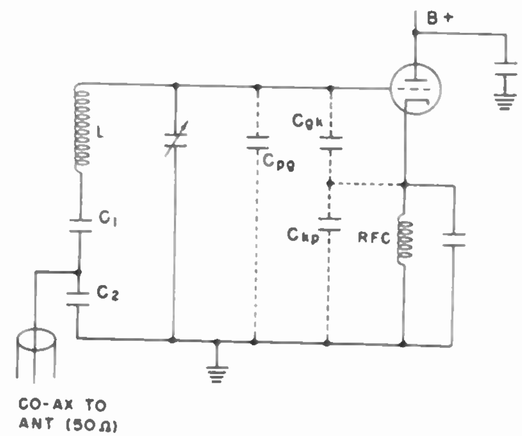


Fig. 7 - Simplified RF oscillator.

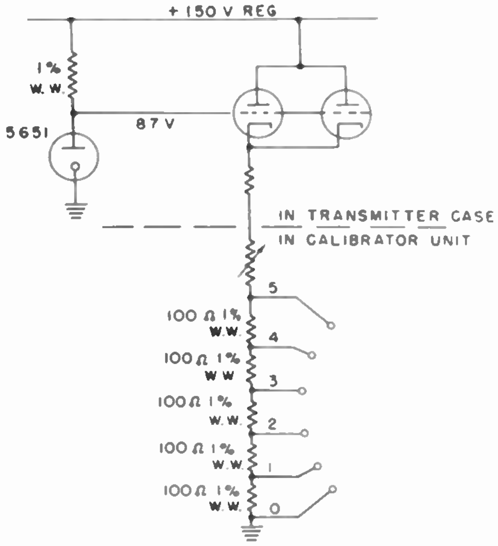


Fig. 8 - Calibration circuitry.

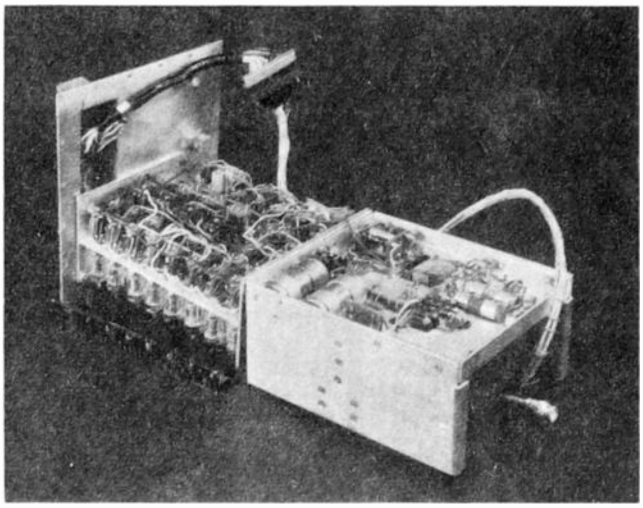


Fig. 10 - Transmitter showing deck construction.

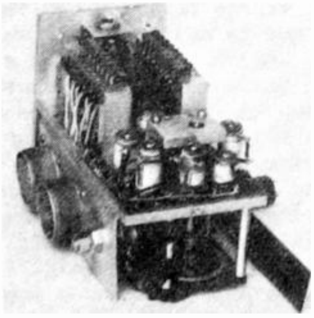


Fig. 9 - Calibrator.

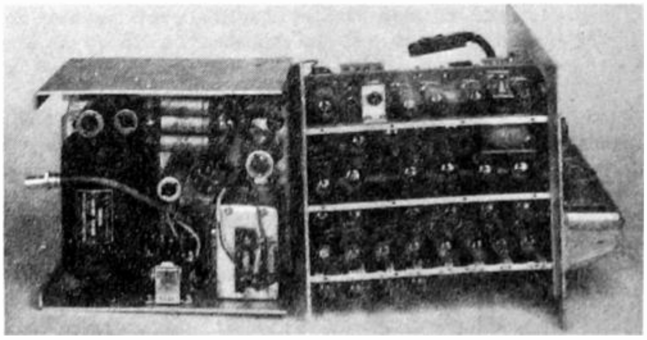


Fig. 11 - Transmitter showing T-plate construction.

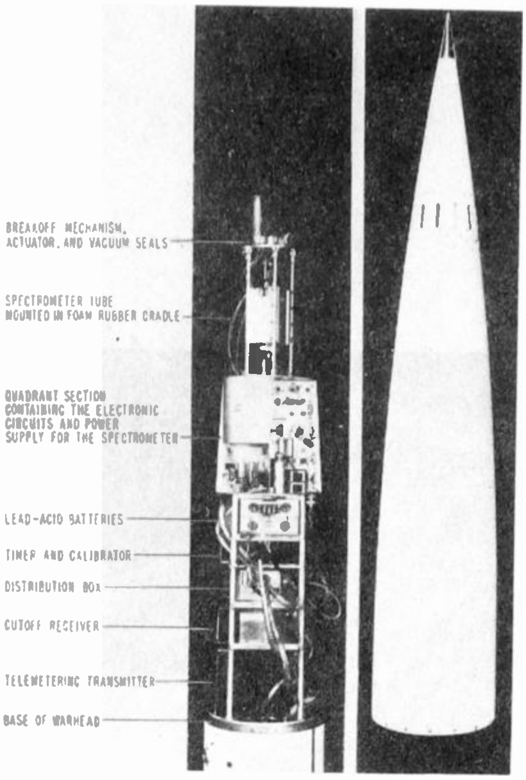


Fig. 12 - Typical Aerobee nose installation.

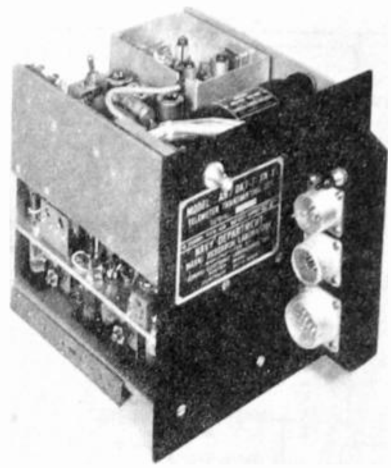


Fig. 13 - AN/DKT-7 (XN-1) transmitter.

CRYSTAL CONTROL LOW DISTORTION FM TELEMETERING TRANSMITTER

R.E. Rawlins
Lockheed Aircraft
Electronic Research Laboratory
Burbank, California

(Abstract was not available at the time of publication.)

A CRYSTAL CONTROL FM TELEMETRY TRANSMITTER

Foster N. Reynolds
Eastern Representative
The Ralph M. Parsons Company
Pasadena, California

One of the greatest sources of trouble and confusion in present day telemetry systems is caused by carrier drift of the airborne radio frequency transmitter. Reactance modulated units have been used for several years primarily because of the fact they provide a simple and economic method of obtaining relatively linear and wide band FM modulation. It also permits employing a high frequency oscillator as the primary source of RF signal which in turn eliminates the necessity of a large number of stages of multiplication.

This type of modulation is inherently good except that by the very nature of a self-excited oscillator a tendency toward center frequency instability is incurred. It is pointed out that practically all environmental conditions to which such a unit is subjected have at least a second if not a first order effect on the normal operating frequency.

This detrimental characteristic has resulted in a vast amount of engineering in the ground receiving stations in order to obtain a compatible radio link. Such things as automatic frequency control and panoramic adaptors have been the end result. Automatic frequency control in particular not only complicates the circuitry of the ground receiver but introduces a finite amount of distortion and intermodulation.

Present restrictions on the frequency spectrum normally employed by telemetering systems have in the past necessitated either time sharing of operations or the reduction of the number of tests made at a particular location. It has always been assumed that the maximum bandwidth required for a normal telemetry RF channel is 250 kilocycles. The drift of the self-excited type of transmitter increases this by an appreciable factor since, if overlapping and interference are to be eliminated, the maximum possible drift must also be taken into account when units are employed in multiple operation. Servicing and operation of such a transmitter are also a problem due to the fact that drifts in center frequency during pre-flight calibration require a considerable amount of searching with the radio receiver in order to determine the operational status of the unit.

The development of this new transmitter was undertaken with the thought in mind of attempting to eliminate as many of these undesirable characteristics as possible. The obvious solution to the drift problem is to employ a quartz crystal as the primary frequency determining element. Phase modulation was considered, but it has several distinct disadvantages. The number of oscillator frequency multiplications which are necessary to obtain the desired bandwidth requires an excessive number of vacuum tubes and consequently produces a rather inefficient and complex power consuming device. A second major objection is the modulation frequency response characteristics. Transmitters of this type inherently tend to fall off in response with a given amount of input signal voltage at the lower end of the audio frequency spectrum. This is highly undesirable in the case of telemetry transmitters since several of the subcarrier center frequencies lie between 300 and 1700 cycles per second.

An alternate system of modulation which at first seemed to provide the simple solution to the problem of crystal control and flat modulation characteristics was that of employing an airborne frequency discriminator as a portion of a closed loop compensating feed back system. The discriminator produced an error voltage proportional to the center frequency drift of a normally reactance tube modulated oscillator. This error voltage generated in the discriminator was transmitted through a low pass filter to the supressor grid of the reactance modulator tube. The net result, therefore, is that should the transmitter have a tendency to drift, a correction signal feed back into the oscillator modulator provides a compensating zero frequency correction. The problem then arose of stabilizing the discriminator so that its drift did not produce an effect similar to a normal drift in the transmitter oscillator. This was accomplished by linking the two coils of the discriminator together. This series link had as a portion of its circuit a quartz crystal which acted essentially as a band pass filter resulting in the discriminator being clamped by a very low impedance shunt when the transmitter carrier frequency was equal to that of the crystal frequency. This particular solution proved very effective but due to the complexity of adjustment of the discriminator it was rejected.

The alternate solution resulted in the perfecting of a mixing and discriminating combination whereby the output of a crystal oscillator is mixed with the output of a reactance modulated oscillator. The differential signal is then discriminated and, as in the previous case, a zero frequency correction is fed back to the reactance modulator tube through an RC low pass filter.

Figure 1 is a circuit diagram of the finished unit. The self-excited oscillator is connected as an electron-coupled Hartley oscillator and is adjusted over a frequency range of 26.875 to 29.375 megacycles. This frequency is doubled in the plate circuit of the oscillator

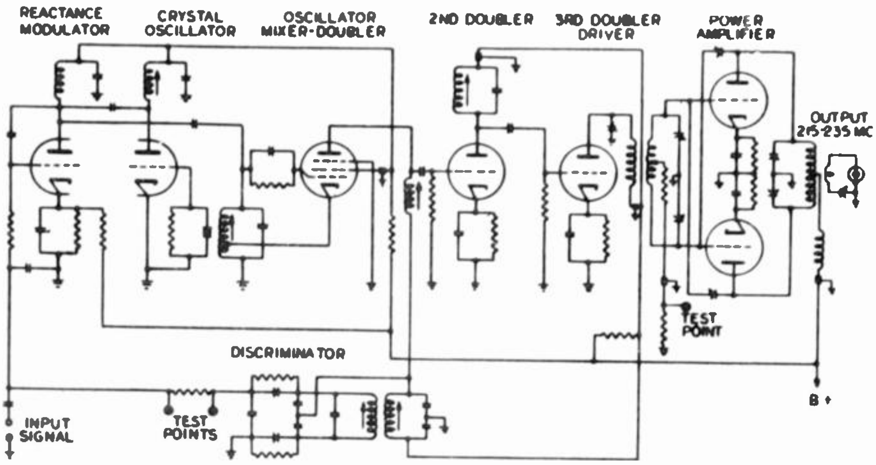


Fig. 1

and then doubled twice again in the second and third doubler driver. The frequency feed to the grids of the power amplifier is thus eight (8) times the frequency of the self-excited oscillator. The final amplifier is a standard push-pull cross neutralized triode amplifier stage. Frequency multiplication was not employed in this final stage so that a maximum power output could be obtained. The output circuit is matched to a 51.5 ohm load, but it may be adjusted with a trimming capacitor located in the transmitter to permit efficient operation with antennas that vary within reasonable amounts from this figure. Pi-section filters are also built into the supply leads to minimize radio frequency radiation from the power cable. Complete transmitter shielding is employed to confine stray fields to the inside of the case. Additional output power in excess of four (4) watts may be obtained by employing a 200 volt plate supply at the expense of some tube life. Five (5) watts have been delivered to a dummy load for periods of several hours with no apparent overheating or tube damage. All of the coils except the final doubler coil and the final amplifier coils are slug tuned. These two are self-supporting and tuned by variable capacitors to provide a maximum efficiency at the higher frequencies.

Modulating signals having a maximum amplitude of 0.25 volts RMS are applied to the grid of the reactance modulator tube through a series RC combination which offers a low impedance path for the modulating signals. As the modulating signal varies in amplitude the reactance modulator acts as a variable capacitor connected across the cathode tank circuit to the oscillator-mixer-doubler. The bypass condenser in the reactance modulator input circuit provides an RF short to ground and this associated network also produces the proper time constant for the frequency control voltage.

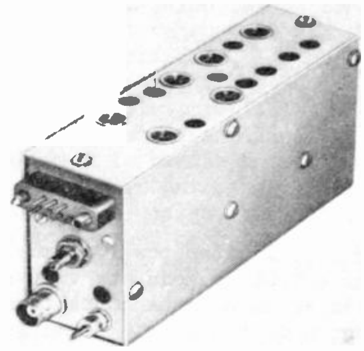
A wide range of plate supply voltage may be used on this transmitter without variation of the carrier center frequency. This is accomplished in part by providing a cathode stabilizing voltage to the reactance modulator tube. It is obtained through a dropping resistor network from the B+ lead.

Frequency stabilization is realized by comparing the output of the variable frequency oscillator after doubling with the output of the crystal controlled oscillator. The resulting difference frequency is fed to a discriminator circuit, the output of which is combined with the modulation signal and returned to the grid of the reactance modulator. An integrating circuit is employed in the output of the discriminator to prevent degeneration of the modulating frequency above 100 cycles per second and to provide a path for the DC control voltage. The Miller circuit of the crystal oscillator uses a third overtone crystal. The output of this oscillator is 5 megacycles less than double the frequency of the self-excited oscillator and mixer. The single tube labeled oscillator mixer-doubler not only acts as an oscillator and doubler, but also as a mixer for its own output and the output of the crystal oscillator. The difference frequency of these two oscillators after doubling of the self-excited frequency is 5 megacycles and it is this signal that is fed to the airborne discriminator. Should the 5 megacycles signal received by the discriminator vary somewhat due to drift in the self-excited oscillator, a positive or negative output depending on the direction of drift will be developed by the discriminator. This voltage then is applied to the reactance tube grid which produces the capacitive reactance necessary to return the unit to the proper operating frequency. The integrating network on the output of the discriminator also prevents the application of a correction signal to the reactance modulator in the presence of a higher frequency input modulation voltage.

Figure 2 is a photograph of the transmitter in its normal operating container. Power connections and primary frequency adjusting elements are all provided at one end of the unit. Other controls which may have to be adjusted if the crystal frequency is changed are available from the top.

Figure 3 is an exploded view of the unit with the outside metal cover removed and the top to which the spring type tube shields are connected placed in a position directly above its normal

Fig. 2



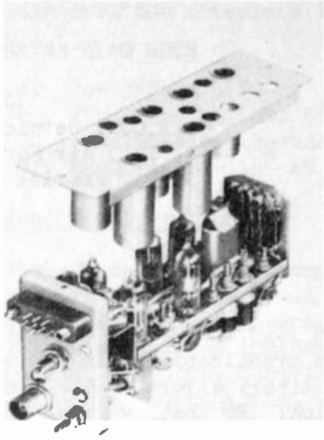


Fig. 3

location. These tube shields were especially designed for this transmitter and they contain a bronze liner placed between the tube envelope and the inner wall of the shield, thus providing excellent mechanical rigidity, as well as very good heat conduction from the tube to the shield. It may be noted from this figure that all tubes are of the replaceable plug-in type and are not soldered into terminal boards, thus assuring a minimum of labor required during maintenance. The crystal is also of the plug-in variety.

The carrier of the transmitter is stable within ± .01% of its center frequency under all environmental conditions and the modulation response is flat within 1.5 db from 100 cycles per second to 100 kilocycles per second. The maximum harmonic distortion introduced under all operating conditions is less than one per cent. The unit employs six subminiature plug-in type vacuum tubes which allow a normal output of 3½ to 4 watts carrier between 215 and 235 megacycles. The volume of the unit is approximately 19 cubic inches and the weight is 14 ounces. The maximum deviation with 0.2 volts RMS input is ± 125 kilocycles with an input impedance of 500,000 ohms shunted by 50 micro-micro farads capacity.

Supply voltages are of standard values. The heater requirement is 25.2 volts DC ± 10% at a current of approximately 300 milliamperes. The B + voltage is normally 180 volts DC at 135 milliamperes, although this may be varied depending on the amount of power output at which the unit is operated. An output impedance of approximately 51.5 ohms through a tunable link is provided, and a maximum of ± 150 kilocycles deviation can be obtained before the linearity specifications are exceeded.

The overall dimensions of the transmitter are five (5) inches by one and a half (1½) inches by two and a half (2½) inches, and it was designed to present a maximum resistance to vibration, shock, acceleration, and temperature changes. Vibration along any axis produces less than one (1) kilocycle deviation in peak carrier frequency; a shock of 60 G for 11 milliseconds along any axis will cause a peak noise of less than 10% of the normal modulation bandwidth; an acceleration of 60 G along any axis results in less than 1% bandwidth variation in the output signal; and temperature variations from - 20° to + 85° Centigrade at altitudes from 0 to 75,000 feet do not affect its operation.

The tube component consists of one (1) Type 6111, one (1) Type 6153, and four (4) Type 5718 subminiature tubes. An additional advantage to this system of modulation and stabilization is that pulse reshaping circuits are not required in the ground station equipment when the transmitter is employed as a portion of a PWM telemetry system.

The engineering on this transmitter was done by Mr. James F. Lawrence presently of The Ralph M. Parsons Company, Pasadena, California. Mr. Lawrence has had experience as a project engineer for Hoffman Laboratories in VHF and UHF communication equipment and with both MGM Studios and the University of Southern California in these same fields. He has also obtained both an AA in engineering and a BE in electrical engineering from the University of Southern California.

HIGH GAIN ANTENNA SYSTEM FOR MULTIPLE OPERATION

James B. Wynn, Jr.
R.C.A. Service Co., Missile Test Project
Air Force Missile Test Center
Patrick Air Force Base, Florida

Introduction

Back in 1942 and 1943 when radio telemetry was first showing the practicability in the structural flighting of military aircraft, the concept was for local operation. By that, we mean that a flight test vehicle operated within the range of a receiving station to record the entire flight program. Perhaps more than one receiving station was utilized to prevent complete loss of data if one station failed, but in any event, we were considering the entire mission being performed within the range of a receiving station regardless of the quantity required for backup purposes. Today, we are considering a test range of at least 1,500 miles in length and under these conditions, it is necessary to have a complete network of receiving stations in order to provide continuous data coverage. The transition made during this period has been from local operation, to a network operation whereby data may be continuously and sequentially collected through many separate stations and furnish accurate correlated data in terms of the function and time.

The need for high gain antenna systems at the receiving stations quickly became apparent. Quite obviously, the basic necessity for a network of stations was based upon the fact the flight tests had extended beyond the possible coverage of a single station. This paper will describe a high gain antenna system that has resulted from the needs of network operation and long range flights.

General Description

The first solution one would propose to extend the transmission range of radio telemetry is rather academic. We should use a frequency suitable for long range transmission, use a high power transmitter and high gain transmitting arrays. The frequency range of radio telemetry we are concerned with at present is in the lower portion of the 200 megacycle region. This band has been assigned by the Defense Department and the FCC so that no choice exists. As to the transmitting components, we should take a closer look at the limitations.

In the applications of radio telemetry many severe restrictions are placed upon the airborne equipment contained in the test aircraft, missile or projectile. In this case, it is the transmitting components that are airborne. The transmitter and associated modulation components must be minaturized, ruggedized and require relatively little power so that they may be installed within a small compartment aboard a missile, survive the high accelerations and vibration, and add little load to a limited power source. This limits the power output to moderate values. In addition, an omnidirectional antenna pattern is desired from an antenna that is not permitted to alter the aerodynamic characteristics of the test vehicle. Needless to say, many of the patterns are poor due to compromise. After considering the limitations imposed upon the airborne transmitting package, it should be apparent that wherever possible, we must consider improvements in the ground receiving stations in order to compensate for (a) relatively low power and (b) poor antenna patterns. Most of the airborne equipment is expended in the tests so that from an economic standpoint, refinements should be incorporated in the receiving stations whenever possible.

It is encouraging to find that the application and facilities associated with the receiving stations do not impose very strigent restrictions upon the receiving components configuration and power requirements. Air-conditioned buildings permit the use of well designed rack mounted equipment with regulated power circuits available wherever required. High gain tracking antennas may be installed for most efficient operation. However, the sites available for the location of receiving stations were limited and from an economical view point minimum coverage overlap should be provided. The actual sites selected in this instance were low flat islands throughout the Bahamas and covering the distance between the central east portion of Florida and Puerto Rico. The elevation of these sites above sea level is relatively small so that little benefit is derived from receiving antenna height.

The most significant and also the most widely used components in the receiving telemetry systems are the antenna system, receiver and magnetic tape recorders. Considerable advancements have been realized during the recent years in regards to both the receivers and magnetic tape recording equipment. Early missile projects started out using a $2\frac{1}{2}$ turn helical antenna with an effective ground plane diameter of approximately one wavelength. The application of helical antennas to the radio telemetry field occured during the later part of the 40's.

Receiving Antennas.

In considering the basic parameters for a radio telemetry receiving antenna, it should be pointed out that flight maneuvers such as high angle dives, rolls, and sometimes the unpredicted tumbling and self destruction often causes the

angle of polarization of the transmitted signal to rotate 360 degrees or more during a short period of time. A very definite requirement thus exists for a receiving antenna which is not sensitive to the angle of the field components and possessing broad band characteristics. In considering the antenna requirements, the selection of the helical antenna was most appropriate. However, the physical construction of a helical antenna makes it both awkward to handle and quite bulky. Not to imply that all other types of receiving antennas are of simple design, but since a helix is comparable to a giant corkscrew which should be supported at only one end, it does present many design problems.

For many of the flight programs, it was calculated that a receiving antenna system having considerably more gain than the relative low gain of about 6 db being realized from most of the helix antennas being used. For this reason, several projects were initiated to provide a receiving antenna system of 30db gain. The purpose of this paper is to present the overall management, objectives, equipment developed and status of a wide band antenna system to provide 30 db gain at an extremely low noise level. Primary breakdown of the project was in two separate parts, namely, development of an improved antenna and development of a wide-band low-noise preamplifier. Refinements that have resulted from this work and will be covered briefly are remote tracking and multicouplers.

Antenna Specifications.

Perhaps it is appropriate at this point to consider in detail the specifications established for the new helical antenna. First, it is worthwhile to mention a few of the practical problems of construction, application and handling that the engineering group had to weigh against theoretical ideals. Under construction we must consider the materials, tooling and assembly techniques available. Application calls for a review of environmental requirements, types and duration of installations, associated equipment characteristics as well as quantities involved. In considering handling, problems such as packageing, rigidity, ease of parts replacement and durability so that breakage would be kept to a minimum, all had to be considered. It was important that an improved antenna be developed, but it was equally important that it be suitable for continued use in the field.

Research into the various helical parameters was conducted on a theoretical basis to establish the limits that could be tolerated in the design criteria. Many tests were conducted, using various diameters of tubing for the helix relative to the problem of supporting the helix assembly. When more than 2 or 3 turns are used for the helix, it is impractical for it to be self supporting. After extensive tests had been conducted, and nearly anyone will agree that antenna tests are elaborate and difficult to conduct, as well as costly, the following salient electrical specifications were established for a helical antenna to operate over the frequency limits of 216 to 236 mcs.

a. Power Gain 10 db or greater
b. VSWR Less than 2:1
c. Polarization Constant over 360°
d. Beam Width 45° at -6db

Mechanical Considerations

Now we should devote a few moments in considering the configuration of the final model that was considered suitable for acceptance evaluation. Several shop prototypes including modifications on a modified prototype were fabricated and field tested before an antenna that was mechanically sound came out of the shop. Recalling that it was previously pointed out that it was desirable to remotely track the antenna, the ground plane was limited to fit in the yoke assembly of a commercially available pedestal made for a six foot parabola. Tests conducted on this size ground plane indicated that a six foot effective diameter was satisfactory. Unlike the old circular solid plate of ¼ inch aluminum that formed the major portion of the old antenna ground plane, the new one consists of wire mesh layed over an aluminum hexagon frame. Wire mesh was obviously selected to reduce the wind loading and weight. The hexagon shape is purely for ease and economy of manufacture because electrically, it is most difficult for anyone to establish much difference in performance from a true circle.

In designing the helical element and its supporting structure, we could truthfully say that things took a turn for the worse. In view of minimum wind resistance and weight, it was desirable to use a small diameter tube stock, but on the other hand, it became difficult to properly support such an assembly mechanically and maintain the desired electrical characteristics. It is worthwhile for us to consider the wind resistance because our antennas are installed throughout an area that hurricanes often cross. The problems of corrosion and leakage paths are very acute along the coastal regions, but even more so in the semitropics where humid salty atmospheres prevail. The helix configuration settled upon, followed along the conventional lines recommended by Mr. Kraus for the beam mode of operation. However, as dictated by our requirements, seven (7) turns were used for the helix from tubing ½ inch in diameter.

Helical Antenna

The helix consists of seven (7) turns of ½ inch copper tubing wound in a clockwise direction looking through the ground plane. Copper tubing was selected because it is possible to sweat it securely to the UG-352/U coaxial connecter, thus providing a good electrical connection and because of the ease in plating copper. The salt atmosphere deteriorates aluminum very rapidly, and it has been difficult at times to prime it properly so that a finish would not easily flake off or scratch through. The silver plated copper is given several spray coats of clear lacquer.

This finish has held up quite well for the past year that we have had them in use.

Since the antenna is working in the axial or beam mode, the diameter of a single turn is 15¼ inches. Spacing between successive turns is 5 3/8 inches so that the overall helical length is 3 feet 9 inches measured from the ground plane surface. Actually, 5 3/8 inches is used for terminating the helix to the coaxial fitting. Four 1 inch polystyrene rods are spaced 90° apart within the helix and forms part of the supporting structure. Two spreaders are used to position the rods uniformly throughout the length of the helix with a pyramid support from the outer spreader to the ground plane. The pitch angle is 12½ degrees. An aluminum 4 inch U beam forms the main support for the antenna ground plane. This beam is formed in the shop out of 0.125 aluminum stock. It's size was determined by two distinct requirements (a) main support for the antenna structure, (b) adequate to house a r-f amplifier directly at the helix termination. The center portion of the beam has been boxed in to form a weatherproof compartment. Four truncated U beam support arms are welded onto the main beam to form the basic structure of the ground plane. All beam members have dimpled, lightning holes wherever practical to reduce the wind loading and weight.

The ground plane actually consists of a wire screen. To be proper, we should call it a wire cloth, since that is the designation given by the manufacturers, Kentucky Metal Products Company. This material may be procured in rolls 6 feet wide which avoids splicing problems. The wire stock used is 0.063 inches diameter with ½ inch openings. Smaller wire diameter was tried, but it would easily wrinkle so the larger wire size had to be selected. This wire cloth weighs 0.59 lbs per square yard, but is desirable over aluminum since it can be sweated directly to the coaxial connector and provide an ideal electrical connection that will not deteriorate.

Perhaps some of you have wondered why a connector like the UG-352/U was selected when it is made to be used with the 7/8 inch coax. Actually the fitting is modified so that the small ¼ inch UG-58/U coax connects to the fitting by two lugs instead of the standard mating part. A lot of the losses in the antenna systems have occured near points of termination where leakage paths build-up due to salt deposits or else the material absorbs moisture. This particular fitting was found to have the greatest surface distance between the inner and outer connecting surfaces and since the material used is teflon, it proved to be the best standard piece of hardware easily adaptable for this use.

Antenna Evaluation.

The electrical characteristics of the new seven (7) turn helical antenna were established by an independent laboratory. The power gain above an isotropic radiator ranged from 11.4 db to 12.0 db in terms of average measured gain. With a broadband antenna the gain is somewhat meaningless without frequency and for this application polarization must also be considered so it is well to look in some detail to a few tabulated figures showing more concisely the gain data.

Polarization	Freq.(MGS)	Gain (DB)		Gain(DB) Average
		E	E	
	216	12.5	9.8	11.4
	226	13.1	10.1	11.8
	236	13.4	10.3	12.0

As you might gather from the foregoing information regarding the frequency characteristics, it was roughly shown that the sensitivity to an illuminating angle of polarization remained fairly constant between the limits of verticle to horizontal wavefront. The circularity of polarization was plotted for the principal axes of the polarization ellipse and tabulated in terms of the ratio in decibels. The axial ratio remained approximately 2.75 db throughout the band.

The nominal magnitude of the impedance of the helical antenna under investigation is approximately 150. Actually, the antenna exhibited a minimum VSWR of 3.1 to 1 over the frequency band. This made it necessary for the fabrication of an impedance transformation section. By the use of a 6.5 inch length of RG-63/U coaxial cable, which exhibits a characteristic impedance of 125 ohms, a satisfactory impedance transformer was fabricated. Between the frequency range of 216 to 235 mcs a maximum VSWR of 1.24 to 1 was achieved. This low standing wave ratio is important for the typical long transmission line feeding a receiver, but an interesting fact in this regard will be pointed out later when we discuss the r-f amplifier.

The directivity of the seven (7) turn helical has been increased or conversely for a comparison with the old 2½ turn helical antenna, the beam width has been decreased from about 60 to 70 degrees to between 42 to 56 degrees at the half power level (-3db).

The seven (7) turn helical antenna was desired for operation over a wider frequency range and this data on the new antenna represents the frequency range of 216 to 236 mcs and throughout the full range of 360 degrees of polarization. A remaining factor that doesn't have too much practical consideration in this particular application, but is worth mentioning, is front-toback ratio. Antenna radiation patterns plotted during the evaluation showed that the minimum front-to-back ratio measured was approximately 7 db.

Tracking Pedestals.

A Houston-Fearless remote control parabola assembly, model RCP-6 was used as the basic tracking pedestal. Since our own antennas have been fabricated, the term parabola as used by the manufacturer does not apply. The modifications for such will be undertaken shortly.

The system naturally comprises two units, the pedestal for positioning the antenna and a control unit for remotely driving as well as indicating the antenna position. We will only briefly review this position of the overall antenna system, since it is relatively common.

The pedestal is modified in our shop to provide 730 degrees of rotation before contacting the electrical limit switches. A total change of 48 degrees in elevation is possible and since the antennas are relatively broad, no modification was required in this respect. The antenna is positioned both in elevation and azimuth by 1/ 6 HP motors which provide a torque of 10,500 inch pounds @ 1 RPM. The pedestal assembly has been stressed for a wind velocity of 120 MPH.Magnetic brakes on the motors prevent the antenna from coasting so that both azimuth and elevation may be positioned within plus or minus 10 minutes of arc.

A control panel,14inches high is mounted in a standard relay rack down in the telemetry receiver room. Correct antenna position is indicated to an operator by observing the relative signal strength of a receiver attached to the antenna. Both azemuth and elevation position of the antenna is indicated on large eight inch dials by means of selsyns. Pairs of antennas may be controlled from a single panel by parallel operation through relays from a single control panel.

Power required for operation is 115 volts 60 cycles single phase at 10 amps. Total weight including the antenna is about 775 lbs, although the antenna weighs only about 60 lbs. Overall height of the complete assembly is about 9 feet 8 inches.

Considerations for R.F. Amplifier & Multicoupler.

The need for increased sensitivity and multiple operation of several receivers from a single antenna arose simultaneously. For most efficient use it was planned to install the r-f amplifier directly at the base of the antenna as pointed out previously. For this reason, the power supply is a separate piece of rack mounted equipment. On the other hand, the multicoupler is intended for rack installation at all times. Multiple operation of 3 or 4 receivers from a single antenna became mandantory as the number of receiving stations increased at each site.

Description of R.F. Amplifier

The r-f amplifier is a wide-band, low-noise high-gain unit of compact design. The gain is sufficiently high so that the preamplifier may be located a considerable distance from the receiver and still maintain an excellent system sensitivity. Overall gain through the amplifier is approximately 15 db with a minimum bandwidth of 20 megacycles. The noise figure of the r-f amplifier is approximately 2.5 db with a gain of approximately 15 db.

The exceptional performance of the preamplifier is mainly contributed by the Western Electric planar triode, tube type 416-A. This tube possesses excellent low-noise characteristics due to its low transit-time and high transconductance. A 6BQ7A dual-triode functions in a cascode circuit following the 416-A and completes the compliment of the preamplifier.

As common for most low noise circuits, the input circuitry is adjusted to give optimum noise factor rather than a correct impedance match. On the other hand, since the amplifier has practically no transmission line between it and the antenna, even relatively high standing wave ratio's are of no consequence. Likewise, sufficient gain is available so the efficiency may be sacrificed for improved quality.

The preamplifier power supply is mounted on a 3½ inch standard rack panel. Continuous metering of the filament voltage, plate current and plate voltage is provided. Electronic regulation is provided for the plus 250 volts.

Description of the Multicoupler

The multicoupler is intended to be an adjunct item to the r-f amplifier so that it must possess the basic characteristics of the amplifier such as wide-band, low-noise and reasonable gain. Since the preamplifier does drive the multicoupler with a relatively high level signal the low noise consideration is not as serious. It is possible to simultaneously feed up to 4 radio receivers from the multicoupler with any signals within the 215-235 megacycle range. Separate stages are used to drive the receivers so that a high degree of isolation exists.

Application of the multicoupler calls for using it in a rack within the telemetry receiving room. It is a recess mounted self-contained unit. occupying 3 inches of panel space. The chassis is recessed so that patching of coaxial cables may be neatly made. All of the tubes and controls extend from the panel for ease of maintenance. Both input & output impedance is the nominal 50 ohms. Insertion gain from the common input to any output is approximately 10 db over the bandwidth of 20 megacycles at the half-power point. Isolation between any output is 25 db or better to minimize interaction between receivers.

The input of the multicoupler consists of a 6AJ4 in a grounded-grid circuit. Following the common amplifier are four cascode-connected 6BQ7A stages to provide separate outputs. Sub-chassis construction is used for the r-f circuitry.

Conclusion

It is well to point out before going further that the system components realized are the result of projects directly performed or initiated and monitored by the Telemetry Unit of Systems Engineering, Air Force Missile Test Center, Patrick Air Force Base, Florida. Most of the antenna development, manufacturing and remote tracking

tracking techniques are the result of Mr. G. E. Bower and Mr. B. G. Miles of the Telemetry Unit. The final experimental and analytical evaluation of the production prototype helical antenna developed by the Telemetry Unit was conducted under Air Force Contract AF 08(606)-602 by the Microwave Engineering Co., of Los Angeles, California. The design and manufacture of prototype wide band r-f amplifiers and electronic multicouplers was performed under Air Force Contract AF 08(606)-605 by the Applied Science Corporation of Princeton, Princeton, New Jersey. Stimulas for the latter was provided by Mr. Lowell, Naval Research Laboratory, by his work on a wide band low noise amplifier in the 200 mc region and the design and fabrication of an antenna multicoupler for the Navy by the J.P. Seeburg Corporation.

An antenna system has resulted from this program at AFMTC that provides an overall gain of 30 db.

Multiple operation of up to 4 receiving stations is made possible with the use of a multicoupler so that the investment in antenna systems is reduced approximately 75%.

Remote tracking is possible and when combined with multiple operation results in a considerable manpower saving.

Acknowledgment is made to both the Applied Science Corporation of Princeton and their staff for the outstanding and agressive handling of the development of both the r-f amplifier and multicoupler, and Mr. Garth Bower, project engineer at AFMTC.

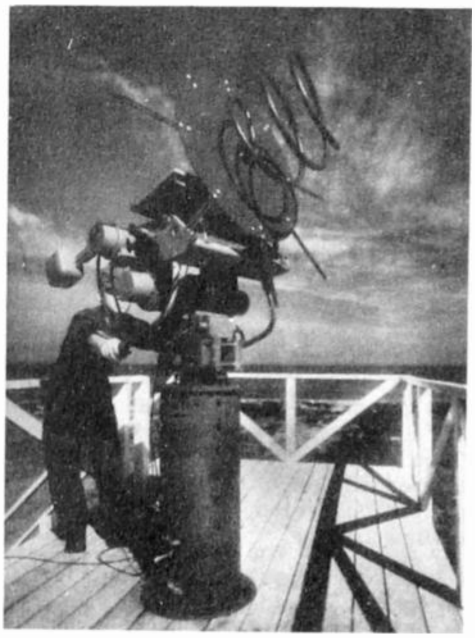


Fig. 1
A view of the old $2\frac{1}{2}$ turn helical antenna mounted on a pedestal for manual tracking. This installation, typical of the early telemetry instrumentation sites throughout the test range requires an operator to manually track each antenna. (U.S. Air Force Photograph.)

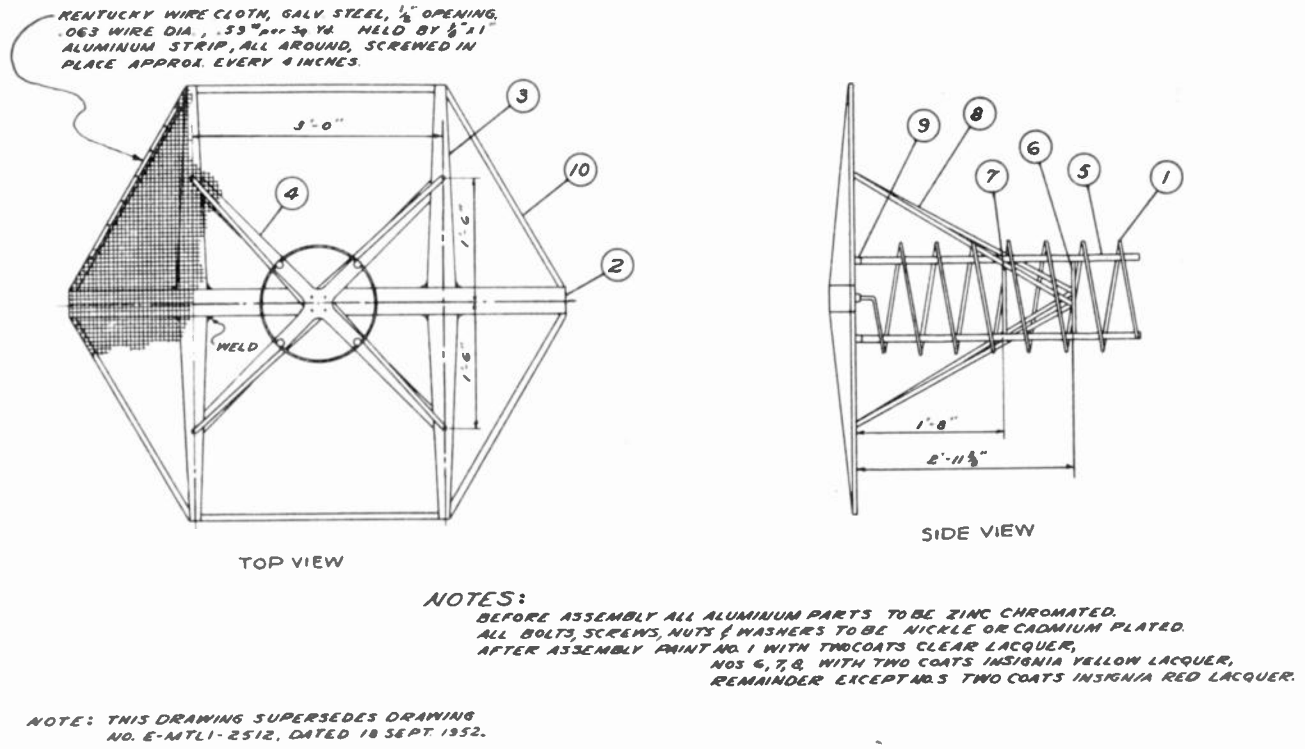


Fig. 2
Assembly drawing showing front and side views of the new 7 turn helical antenna. The helix, constructed of $\frac{3}{4}$-inch copper tubing is supported on 4 polystyrene rods. Pyramid support rods provide high rigidity to the helix assembly.

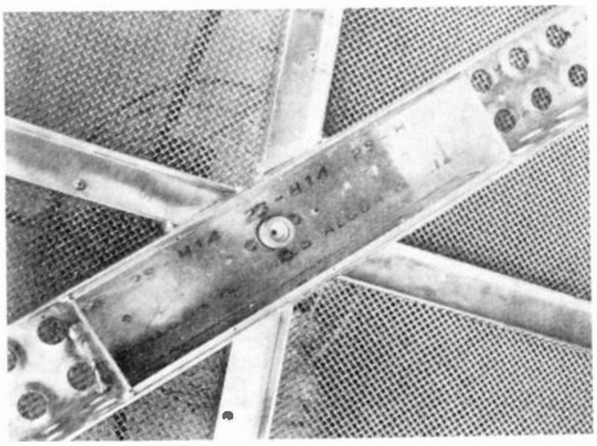


Fig. 3
A close-up view of the main beam of the helical antenna ground plane. Antenna termination is through the UG-352/U coaxial fitting shown in the center. This portion of the main beam is partitioned off to provide a weather proof housing for the low-noise RF amplifier.
(U.S. Air Force Photograph.)

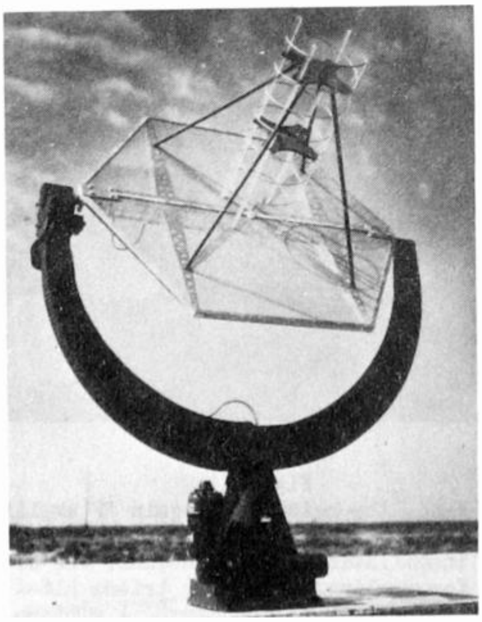


Fig. 4
A view of the new high-gain 7 turn helical antenna and associated remotely controlled pedestal mounted on the roof of one of the instrumentation buildings. Minimum average gain is 11.4 db with 20 mcs bandwidth.
(U.S. Air Force Photograph.)

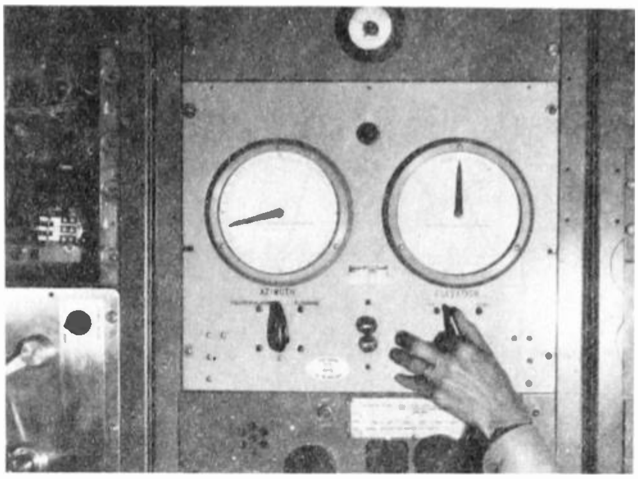


Fig. 5
An operator is shown adjusting the elevation of one of the telemetry antennas at the remote control panel. Both azimuth and elevation position of the antenna is indicated on the large dials by means of selsyns. Correct antenna position is indicated by observing the relative signal strength indicated on the meter mounted above the control panel.
(U.S. Air Force Photograph.)



Fig. 6
The wide-band, low-noise, high-gain RF amplifier is shown being inspected by M/Sgt D.C. Taylor prior to installation on the antenna. The blower required for cooling the planar triode 416-A tube is lying at the rear of the bench. A minimum gain of 15 db over the 216-236 mcs range is possible by use of this amplifier.
(U.S. Air Force Photograph.)

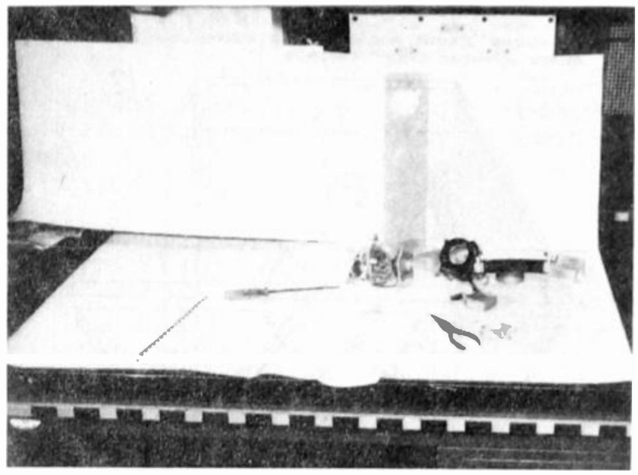


Fig. 7
A view of the RF amplifier removed from the mounting panel and case. Normally, the amplifier is installed directly at the antenna termination. The noise figure of the RF amplifier is approximately 2.5 db.
(U.S. Air Force Photograph.)

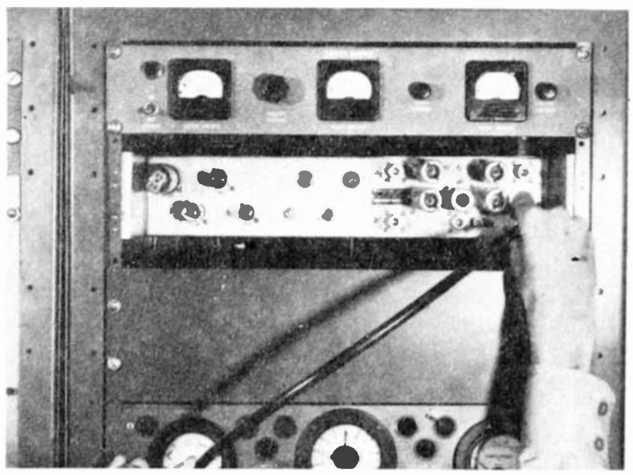


Fig. 8
The recessed, rack mounted multicoupler is shown being set up for operation. It is possible to simultaneously feed up to 4 separate radio receivers from the multicoupler with signals within the 216-236 mcs range with an insertion gain of 10 db. Isolation between any output is 25 db or greater. Mounted above the multicoupler in this illustration is the power supply for the RF amplifier.
(U.S. Air Force Photograph.)

OPERATIONAL ANALYSIS OF TRACK-WHILE-SCAN RADARS

Stephen J. O'Neil
Air Force Cambridge Research Center, ARDC
Cambridge, Mass.

Summary. A technique is presented for evaluating the performance of search radars used with track-while-scan equipment. For radars with fixed beam width and pulse repetition frequency a change in the antenna scan rate changes both detection probability and target motion between looks. Since these changes have opposite effects upon tracking performance, there is an optimum antenna scan rate. Detection probability and possible deviations from predicted straight-line flight motion are used to define the mathematical expectation of tracking error. The tracking error is used as a criterion of tracking performance. It is shown that a minimum value of mathematical expectation of tracking error is obtained if the antenna scan rate is high enough to produce a small number of hits per beam width. A method is presented for determining the width of the tracking gates from the mathematical expectation of tracking error and the number of missed looks beyond which it is no longer considered feasible to continue tracking.

Introduction

An operational analysis may be conducted to determine the most effective manner of utilizing complex equipment. This paper presents a technique for increasing the effectiveness of radar equipment used for Air Traffic Control.

Many authorities[1] share the opinion that an automatic system of Air Traffic Control is desirable to handle the increasing volume of airport traffic. This is particularly true during instrument flight conditions.

Airport radar is used to search the sky about the airport in order to keep the area under continuous surveillance. In some cases the radar also is used to keep track of individual aircraft as they are guided down to the final approach point[2]. As the radar searches the sky, electronic gating techniques are used to examine a particular volume of air space about the aircraft during each radar scan. The tracking gates are repositioned each scan so that the volume of air space to be examined will enclose the predicted position of the aircraft.

Since this problem is relatively new, very little consideration has been given to the most effective use of the radar in conjunction with the tracking system. One might ask, for example, whether the antenna rotation rate influences the operation of the tracking system. A low antenna scan rate results in a large number of "hits per beam width", i.e., target illuminations as the radar beam sweeps through the target position. This increases the probability of detection. With a high antenna scan rate on the other hand the period between radar looks at the aircraft is reduced and the aircraft motion between looks is also reduced. This reduces the possibility that the aircraft might move far enough between looks to get completely outside of the volume of air space positioned by the tracking gates. It is reasonable to conclude that some optimum antenna scan rate will result in the most favorable tracking conditions.

If the target signal is too weak to be detected during some of the radar looks, the aircraft may move a considerable distance between successful looks. If the aircraft performs some sort of maneuver, while the predicted position for the volume of air space to be examined moves in a straight line at constant velocity, the possibility that the aircraft may move entirely out of this volume increases. This is illustrated in Fig. 1.

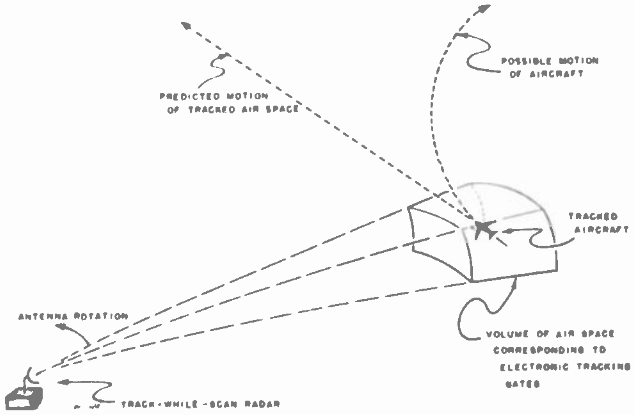


Fig. 1 - The Track-While-Scan Problem

If the target can move outside of the volume of air space, wider electronic gates, which result in larger volumes of air space, may be necessary. This procedure is undesirable since it reduces the tracking resolution possible. Alternatively, narrow gates and occasional human intervention to reposition the air space about the aircraft can be employed. This procedure is also undesirable since the purpose of the track-while-scan equipment is to reduce insofar as possible the need for human intervention in the control center.

It is desirable therefore to establish some basis for determining the optimum antenna scan rate and the optimum tracking gate width.

Problem

The problem studied in this paper is the determination of the optimum search radar antenna scan rate for use with a track-while-scan system.

The tracking system considered predicts future positions for the aircraft by extrapolating the present motion along a straight line at constant velocity.

The problem is treated in two dimensions only. Altitude is not considered in this paper.

Method of Approach to Problem

The tracking error is defined for the purposes of this paper as the difference between the predicted aircraft position, determined by the center of the air space corresponding to the tracking gates, and the actual position of the aircraft. The tracking error is used as a criterion of tracking performance.

Straight-line, constant-velocity extrapolation of measured aircraft position is assumed in order to adjust the tracking gates and establish the predicted position of the aircraft.

In order to test the safe operation of the system, aircraft maneuvers with the most severe initial conditions are considered, and the actual positions of the aircraft are calculated. Two types of aircraft maneuvers are considered. The first is a slow, circular turn with a low wing-load factor such as is typical of aircraft flying some portions of an airport traffic pattern. The second maneuver is a sudden change in course. The latter is a more extreme maneuver from the viewpoint of tracking system performance.

Two types of radar receiver systems will be considered. In the first type provision is made for storage and integration of signal and noise voltages as a function of position in space. In the second type of receiver system only one signal return per beam width is used, and any return successfully received will yield sufficient information for the tracking computer.

It is assumed that while the aircraft is maneuvering, there is a possibility that the target signal is too weak to be detected during some of the radar looks. In order to take this factor into consideration the probability of successful radar detection during a single illumination of the target will be specified.

The probability of detection during each look after a successful look at the target is calculated as a function of the number of hits per beam width.

The detection probabilities and the tracking errors for the two maneuvers considered are used to calculate the mathematical expectation of tracking error which is plotted as a function of the number of hits per beam width. From these plots the number of hits per beam width which results in minimum mathematical expectation of tracking error is determined. The corresponding antenna scan rate is the optimum scan rate for the conditions assumed.

Aircraft Maneuvers

Circular Turns

Fig. 2 illustrates some unfavorable conditions which may exist during a circular maneuver.

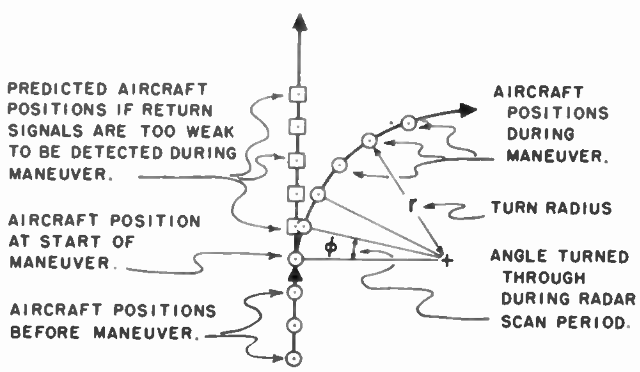


Fig. 2 - Possible Deviation between Actual and Predicted Aircraft Positions During a Circular Turn Maneuver.

In Fig. 2 if ϕ is a small angle, the tracking error during one scan period is given approximately by:

$$\mathcal{E}_1 \cong r\ (1-\cos\ \phi) \qquad (1)$$

where $\mathcal{E}_1$ = tracking error during one scan period,

r = aircraft turn radius,

ϕ = aircraft turn angle during one scan period.

From the series expansion for cos ϕ,

$$\cos\ \phi \cong 1 - \frac{\phi^2}{2} \quad . \qquad (2)$$

Substituting Equation (2) into Equation (1),

$$\mathcal{E}_1 \cong r\ (\frac{\phi^2}{2}) \quad . \qquad (3)$$

As shown in Appendix I the aircraft turn radius for a true-banked turn is a function of its forward velocity and wing load factor[3]:

$$r = \frac{V^2}{g\sqrt{G^2 - 1}} \quad , \qquad (4)$$

where V = aircraft forward velocity

g = acceleration of gravity

G = wing-load factor.

The angular turn or heading change during one scan period is

$$\phi = \frac{V\Delta t}{r} \quad , \qquad (5)$$

where Δt = radar scan period.

The scan period is inversely related to scan rate:

$$\Delta t = \frac{1}{\omega} \quad , \qquad (6)$$

where ω = radar antenna rotation rate.

If ϕ and r are eliminated from Equations (3), (4), and (5), the following result is obtained for $\mathcal{E}_1$:

$$\mathcal{E}_1 = \frac{1}{2}\ (\Delta t)^2\quad g\sqrt{G^2 - 1} \quad . \qquad (7)$$

Equation (7) shows that the tracking error per scan for a circular turn is a function of both the radar antenna scan rate and the aircraft wing-load factor.

If the return signal is not detected for a number of scan periods after the maneuver starts, the tracking error is given by

$$\mathcal{E}_k = k^2\mathcal{E}_1 \quad , \qquad (8)$$

where $\mathcal{E}_k$ = tracking error during k scan periods,

k = the number of radar looks at the aircraft after the last successful look.

Equation (8) is a good approximation if the total angular turn of the aircraft during k scan periods is a small angle.

Sudden Changes in Course

In the interests of safety, situations will be considered in which the aircraft does not behave as predicted. In these situations it will be assumed that the aircraft suddenly changes its course. A sudden change in wind or an extremely rapid large-angle turn could produce this effect. A sudden change in course of sixty degrees probably represents the worst situation of this nature which could be expected. Under these conditions the tracking error per scan is given by

$$\mathcal{E}_1 = V\Delta t \quad . \qquad (9)$$

In this case the tracking error per scan is a function of both aircraft velocity and radar scan rate.

If the return signal is not detected for (k-1) radar looks after the start of the maneuver, the tracking error after k scan periods will be

$$\mathcal{E}_k = k\mathcal{E}_1 \quad . \qquad (10)$$

Radar Characteristics

Probability that a Signal Will Exceed the Threshold Level (p)

Pulsed radars only are considered in this paper. Clutter and clutter-reducing methods such as Moving Target Indicators are not considered. It is assumed that the antenna gain is constant over the antenna beam width. The effect of the antenna beam pattern is discussed in Appendix II.

The number of radar hits per beam width is given by

$$n = (PRF)(BW)\Delta t \quad , \qquad (11)$$

where n = the number of hits per beamwidth,

PRF = radar pulse repetition frequency,

BW = radar beam width,

Δt = radar scan period.

If the beam width and pulse repetition frequency are fixed, then the number of hits per beam width is proportional to the duration of the scan period.

The probability of radar detection as a function of range has been determined by several authors[4,5,6,7]. Using the methods developed in references 4, 5, and 6, the detection probability for a single return signal has been calculated as a function of range for various values of the false alarm time. These probabilities are plotted in Fig. 3.

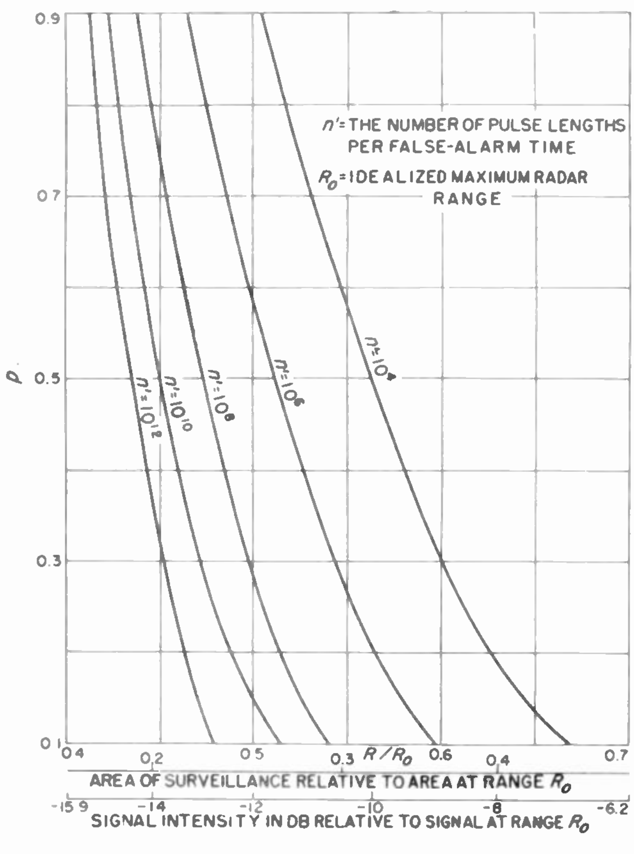


Fig. 3 - Probability (p) that a Return Signal Will Exceed the Threshold Level as a Function of Range (R)

The false alarm time is the average length of time between noise pulses which exceed the threshold level; it is assumed that the threshold level is adjusted to a value such that the probability of a noise pulse occurring during the false alarm time is one-half.

Probability that at Least One Signal per Look Exceeds the Threshold Level without Integration (P_1)

If at least one signal per beam width exceeds the threshold level, the tracking computer will operate properly. It will be assumed that the various signal returns during a particular radar look can be treated as independent events with the same probability of exceeding the threshold. Then the probability that at least one signal exceeds the threshold is given by

$$P_1 = 1 - q^n \quad , \qquad (12)$$

where P_1 = probability of detection during one look,

q = probability of failure to detect a return signal during one hit (target illumination).

$$q = 1 - p \quad , \qquad (13)$$

where p = probability of detection during one hit.

P_1 is plotted as a function of n for various values of p in Fig. 4.

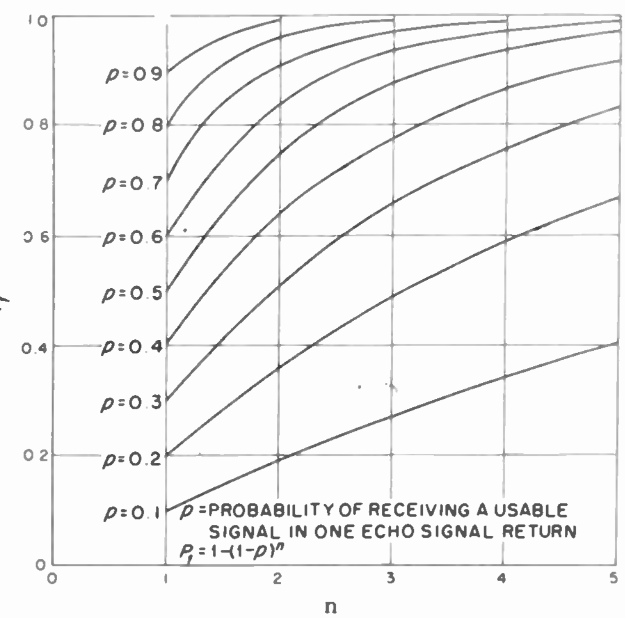


Fig. 4 - Probability (P_1) of Receiving at Least One Signal per Look Exceeding the Threshold Without Integration as a Function of the Number of Hits per Beam Width

Probability that an Integrated Signal Exceeds the Threshold Level in One Look (P_1)

The detection probability for an integrated signal has been calculated as a function of the number of hits per beam width for various values of p; the results are plotted in Fig. 5.

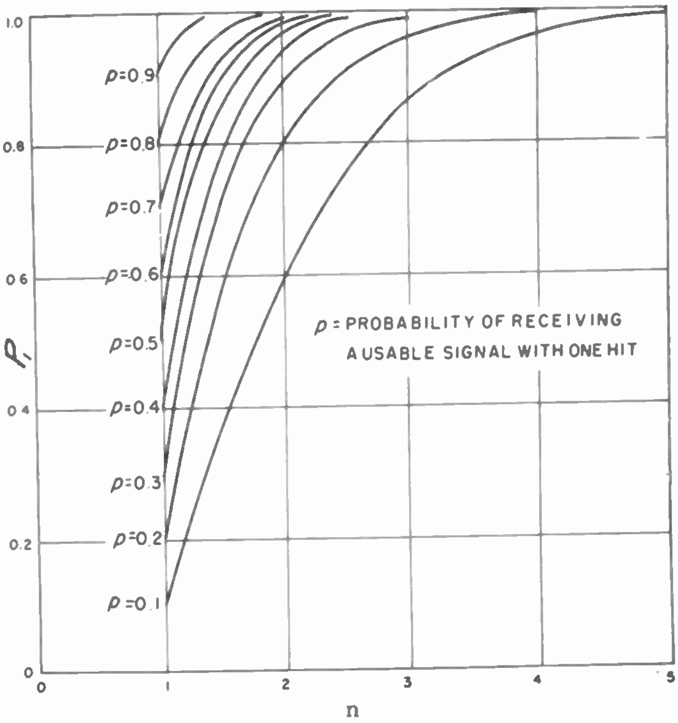


Fig. 5 - Probability (P_1) of Receiving a Signal Exceeding the Threshold with Integration of All Hits per Look (Threshold Signal Level is Determined by n; False Alarm Time = $(10)^6$ Pulse Lengths)

The methods and results of references 4 and 5 have been used in these calculations. The results shown in Fig. 5 are subject to the following conditions:

(1) The number of pulses integrated is equal to the number of hits per beam width.

(2) All of the pulses integrated contain the target signal.

(3) The threshold level is set at a value which corresponds to a false alarm time of $(10)^6$ pulse lengths.

Probability that the Signal will Exceed the Threshold During the k_{th} Look (P_k)

The probability that the signal will exceed the threshold during the k_{th} radar look is given by the product of the probability that no signals exceed the threshold for (k-1) looks and the probability that the signal exceeds the threshold during one look

$$P_k = P_1 \, (1-P_1)^{k-1} \quad . \tag{14}$$

This equation holds for both integrated and non-integrated signals. Without integration P_1 is calculated by means of Equation (12) or obtained from Fig. 4. For integrated signals P_1 is obtained from Fig. 5.

Mathematical Expectation of Tracking Error

The mathematical expectation[8] of tracking error during any radar look is given by

$$\mathcal{E} = \sum_{k=1}^{\infty} P_k \mathcal{E}_k \tag{15}$$

This equation applies if the tracking gates are so wide that an infinite number of return signals can be missed without the aircraft moving outside of the gates. There are practical limits, however to the gate width and to the number of scan periods between useful signals. The following modification of Equation (15) may be used:

$$\mathcal{E} = \frac{\sum_{k=1}^{m} P_k \mathcal{E}_k}{\sum_{k=1}^{m} P_k} \tag{16}$$

where m is the number of consecutive looks without a return signal beyond which it is not considered feasible to continue tracking.

As shown in Appendices III and IV the use of Equation (15) leads to certain simplified expressions. Equation (15) is approximately true and can be applied if $P_1 \leq 0.6$ and $m \geq 5$. Under these conditions the error involved is not greater than 12 per cent.

In this paper Equation (15) is used to study the optimum scan rate problem, and Equation (16) is used with a value of m=5 to determine the width of the tracking gates.

Mathematical Expectation of Tracking Error During Circular Turn

If $\mathcal{E}_k$ from Equation (8) and P_k from Equation (14) are substituted into Equation (15), the following expression is obtained:

$$\mathcal{E} = \mathcal{E}_1 P_1 \sum_{k=1}^{\infty} k^2 (1-P_1)^{k-1} \qquad (17)$$

As shown in Appendix III, for m=∞ :

$$\mathcal{E} = \frac{2 - P_1}{P_1^2} \mathcal{E}_1 \qquad (18)$$

Equation (7) shows that for a circular turn $\mathcal{E}_1$ is proportional to $(\Delta t)^2$ and Equation (11) shows that Δt is proportional to n. Then from Equation (18),

$$\mathcal{E} = \frac{2 - P_1}{P_1^2} an^2 \qquad (19)$$

where a is a function of G:

$$a = \frac{g\sqrt{G^2-1}}{2(PRF)^2(BW)^2} \qquad (20)$$

In the same manner Equation (16) becomes

$$\mathcal{E} = \frac{an^2 \sum_{k=1}^{m} k^2 (1-P_1)^{k-1}}{\sum_{k=1}^{m} (1-P_1)^{k-1}} \qquad (21)$$

Using a fixed value of a and several values of p, Equations (19) and (21) have been evaluated for return signals with and without integration. The resulting values of $\frac{\mathcal{E}}{a}$ are plotted versus n in Figs. 6 and 7.

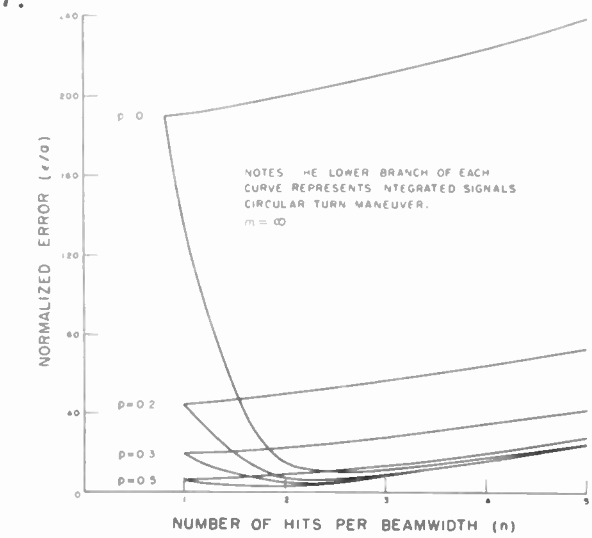


Fig. 6 - Mathematical Expectation of Tracking Error for Circular Turns, m=∞

Mathematical Expectation of Tracking Error During Sudden Change in Course

If $\mathcal{E}_k$ from Equation (10) and P_k from Equation (14) are substituted into Equation (15), the following result is obtained:

$$\mathcal{E} = \mathcal{E}_1 P_1 \sum_{k=1}^{\infty} k (1-P_1)^{k-1} \qquad (22)$$

As shown in Appendix IV, for m= ∞,

$$\mathcal{E} = \frac{\mathcal{E}_1}{P_1} \qquad (23)$$

From Equations (9) and (11),

$$\mathcal{E}_1 = bn \qquad (24)$$

where b is a function of V:

$$b = \frac{V}{(PRF)(BW)} \qquad (25)$$

From Equations (23) and (24),

$$\mathcal{E} = \frac{bn}{P_1} . \qquad (26)$$

In the same manner Equation (16) becomes

$$\mathcal{E} = \frac{bn \sum_{k=1}^{m} k (1-P_1)^{k-1}}{\sum_{k=1}^{m} (1-P_1)^{k-1}} \qquad (27)$$

Using a fixed value of b and several values of p, Equations (26) and (27) have been evaluated for return signals with and without integration. The resulting values of $\frac{\mathcal{E}}{b}$ are plotted versus n in Figs. 8 and 9.

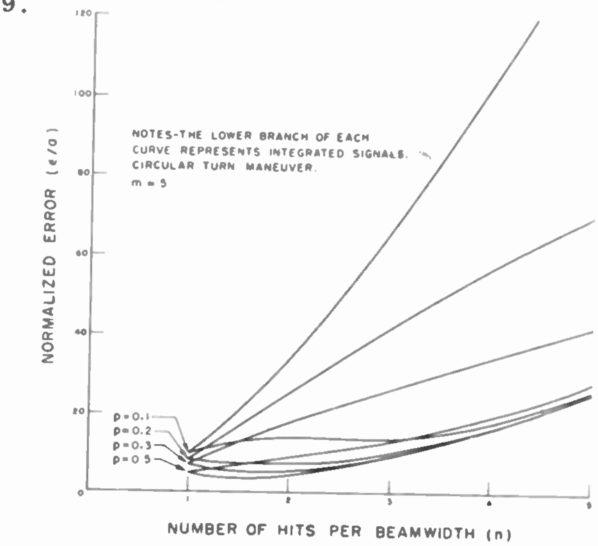


Fig. 7 - Mathematical Expectation of Tracking Error for Circular Turns, m=5

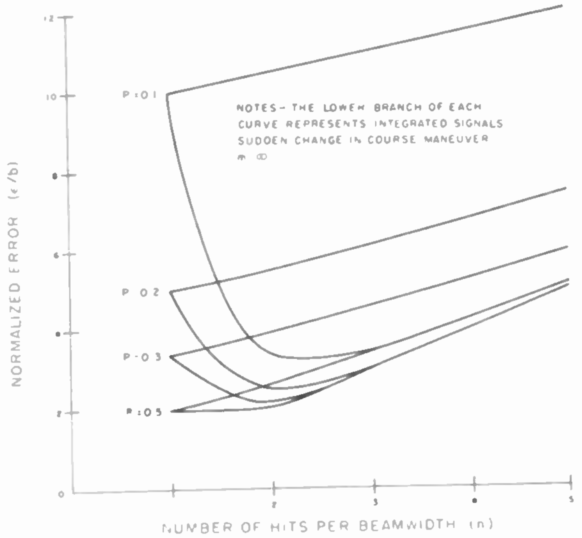


Fig. 8 - Mathematical Expectation of Tracking Error for Sudden Changes in Course, m=∞

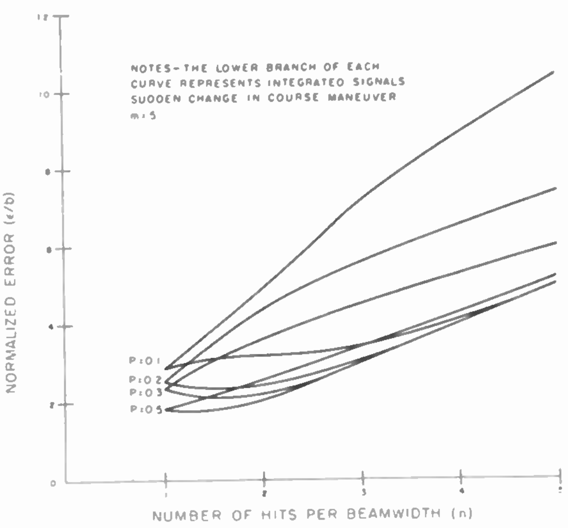


Fig. 9 - Mathematical Expectation of Tracking Error for Sudden Changes in Course, m=5

Optimum Radar Scan Rate

The optimum radar scan rate is that rate at which the tracking error is a minimum. The corresponding value of n, the number of hits per beam width, is determined in this paper. Since the pulse repetition frequency and the antenna beam width are fixed, the scan rate is uniquely determined by the following expression:

$$n = \frac{(PRF)(BW)}{\omega} \qquad (28)$$

From Figs. 6 and 8 it is observed that for systems without signal integration the minimum tracking error occurs if there is one hit per beam width. For systems with signal integration the minimum tracking error occurs if there are two or three hits per beam width.

It can also be observed in Figs. 6 and 8 that the improvement in tracking accuracy achieved through signal integration is most noticeable for low values of p. Since this corresponds to a long range, it may be said that signal integration is of most value in the fringe areas of the radar detection pattern.

Figures 6 and 8 are drawn for a value of m=∞ which means that a very large number of missed looks can be tolerated. Practically the gate width should be a function of the expected tracking error.

The following procedure might be used to determine the proper value of m. Starting with the expected range performance from which p can be obtained, the optimum number of hits per beam width and the resulting value of $\mathcal{E}/a$ or $\mathcal{E}/b$ can be determined from Fig. 6 or Fig. 8. The values of p and n can be used with Fig. 4 or Fig. 5 to determine P_1. Then P_1 and $\mathcal{E}/a$ or $\mathcal{E}/b$ can be used with Equation (21) or (27) to determine the proper value of m.

As an alternative procedure the values of p and m may both be specified. Then $\mathcal{E}$ can be calculated from Equations (20) and (21) or (25) and (27) as illustrated for m=5 in Fig. 7 or Fig. 9. For n=5 the optimum number of hits per beam width and the corresponding value of $\mathcal{E}/a$ or $\mathcal{E}/b$ can be determined from Fig. 7 or Fig. 9.

In either case the width of the tracking gates can be determined from the value of $\mathcal{E}$. The latter procedure has been used in the examples solved in the next section.

Tracking Gate Width

An estimate of the width required in the tracking gates with a given value of m can be obtained from the mathematical expectation of tracking error in m scans. The gate widths should be set equal to approximately twice the value of $\mathcal{E}$.

The following example will be solved for the tracking gate width with m=5 using the two types of maneuvers previously described and considering systems with and without signal integration:

Target Data:

Speed - 180 mph

Load Factor - 1.03

Radar Data:

Scan Rate - 15 rpm

Azimuth Beam Width - 0.8 degree

Pulse Repetition Frequency - 360 pulses per second

The target range and cross section are assumed to have such values that the probability of detection in one hit, p = 0.1.

Solution

1. Circular Turn - No Signal Integration. In terms of the units given above, the number of hits per beam width is given by

$$n = \frac{1}{6}\,\frac{(BW)(PRF)}{\omega}$$

$$= \frac{1}{6}\,\frac{(0.8)(360)}{15} = 3.2$$

Since only an integral number of hits has meaning, this result must be interpreted to mean n=3.

In terms of the units given, Equation (20) is evaluated as

$$a = \frac{2.09(10)^6\sqrt{G^2-1}}{(PRF)^2(BW)^2}$$

$$= \frac{(2.09)(10)^6\sqrt{(1.03)^2-1}}{(360)^2(0.8)^2}$$

$$= 6.2 \text{ feet}$$

From Fig. 7, for n=3, p=0.1 and no integration,

$$\mathcal{E}/a = 67.5$$

Therefore

$$\mathcal{E} = (67.5)(6.2) = 420 \text{ feet.}$$

The tracking gate should be about 900 feet wide.

2. Sudden Course Change - No Signal Integration. In terms of the units given, Equation (25) is evaluated as

$$b = \frac{(264)(V)}{(PRF)(BW)}$$

$$= \frac{(264)(180)}{(360)(0.8)} \quad .$$

$$= 165 \text{ feet} \quad .$$

From Fig. 9, for n=3, p=0.1, and no signal integration,

$$\mathcal{E}/b = 7.2$$

Therefore

$$\mathcal{E} = (7.2)(165) = 1200 \text{ feet.}$$

The tracking gate should be 2400 feet wide.

These results may be interpreted as follows. Without signal integration, an aircraft maneuvering as expected during Air Traffic Control requires a tracking gate 900 feet wide. If the aircraft maneuvers in an unexpected manner, a tracking gate 2400 feet wide may be necessary. A good compromise between target resolution and tracking reliability will probably be achieved by using a tracking gate about 2000 feet wide.

3. Circular Turn - Signal Integration. From Fig. 7, for n=3, p=0.1 and signal integration, $\mathcal{E}/a=14$.

Therefore

$$\mathcal{E} = (14)(6.2) = 90 \text{ feet.}$$

It might seem from this result that a tracking gate 200 feet wide could be used. However, the tracking gate should be larger than the equivalent pulse length. A 1/2 microsecond pulse corresponds to a range interval of 250 feet. A tracking gate 450 feet wide is a better approximation in this case.

4. Sudden Course Change - Signal Integration. From Fig. 9, for n=3, p=0.1 and signal integration, $\mathcal{E}/b=3.5$.

Therefore,

$$\mathcal{E} = (3.5)(165) = 600 \text{ feet.}$$

The tracking gate should be 1200 feet wide.

These results indicate that with signal integration a good compromise between target resolution and tracking reliability will probably be achieved by using a tracking gate about 1000 feet wide. (In practice with a fixed azimuth beam width it may be necessary to use wider gates to encompass the signal from long-range targets).

In this problem signal integration has produced greater target resolution.

Discussion

It is assumed in this paper that the tracking gates are positioned by extrapolating the motion of the aircraft along a straight line. Using straight-line extrapolation two aircraft position measurements and a first order finite difference are necessary to compute a predicted aircraft position for tracking gate adjustment. A more refined method of course extrapolation using higher order differences may be used to position the tracking gates. Although this paper has not mentioned the problem of data accuracy, it is felt that the use of higher order differences should be preceded by a study of the limitations of data accuracy, the possibility of data smoothing, etc.

Straight-line extrapolation has advantages over linear range and azimuth angle extrapolation. Straight-line motion is the median of all possible changes in direction. Therefore this type of tracking results in the least value of the maximum tracking error in one scan.

In the case of linear range tracking, if the aircraft is flying a course normal to the radar bearing, two equal values of range might be measured, and a circular path about the radar inferred. In this case if the aircraft turned away from the radar, the maximum tracking error in one scan would have a greater value than for straight-line tracking. The results of this paper, however, may be applied to the case of linear range tracking. The circular-turn maneuver with a slightly greater value of error per scan may be employed, or the results for the sudden-change-in-course maneuver may be used.

The expected azimuth tracking errors under the most unfavorable initial conditions can be obtained from the expected range errors by applying the following formula:

$$\Theta = \frac{\varepsilon_r}{R} , \qquad (29)$$

where Θ = mathematical expectation of azimuth tracking error,

ε_r = mathematical expectation of range tracking error,

R = aircraft range.

Since the ratio of azimuth beam width to the number of hits per beam width represents the limit of accuracy of the azimuth data, Equation (29) is applicable only if Θ exceeds this figure.

The results presented in this paper assume some ideal conditions for integration. In particular it is assumed that the threshold level is set for a false alarm time which is determined statistically by integrating n sweeps, all of which contain the return signal. It is impossible, however, to tell beforehand on which sweeps the signal will occur, so in general not all of the sweeps in any integrated group will contain the signal. Therefore, the results presented in Figs. 6 to 9 for integrated signals represent minimum attainable values and the difference shown between integrated and non-integrated signals will probably not often be realized in practice.

The results are also based on the assumption that the antenna gain is constant over the beam width. In general this will not be true. The antenna gain and consequently the probability, p, that a return signal exceeds the threshold level, vary over different parts of the beam width. Because of the rotation of the antenna, the return signal from a target will be received when the antenna is pointed in a slightly different direction than when the signal was transmitted. Therefore, p also will vary with the antenna rotation speed. The average value of p over the beam width will decrease as the antenna rotation speed increases or the number of hits per beam width decreases. If p is very low for small values of n, the minimum tracking error in practice may not occur at n=1. In general, however, the results show that the antenna scan rate should be increased to the point where the probability, p, that a return signal exceeds the threshold begins to decrease rapidly.

The results presented in this paper show that in general an antenna scan rate which is high enough to produce a small number of hits per beam width is desirable for accurate tracking. The results also show that some improvement in tracking accuracy is possible with signal integration particularly at long ranges. A lower antenna scan rate and a larger number of hits per beam width are desirable with signal integration.

The difference between the results for the mathematical expectation of tracking error for $m=\infty$ and $m=5$, illustrated in Figs. 6, 7, 8, and 9 may be interpreted as follows. The curves for $m=\infty$ illustrate the tracking error (and necessary gate width) for purely automatic tracking at a given range. The curves for $m=5$ illustrate the tracking error (and necessary gate width) for the same system at the same range when it is successfully tracking with no more than 5 consecutive missed looks. Smaller tracking errors are

obtained in the latter case; however, in order to maintain continuous tracking in the latter case, occasional manual intervention may be necessary.

Conclusions

This paper has presented a theoretical method of evaluating a complete track-while-scan system. The mathematical expectation of tracking error has been computed and used for this purpose.

It has been shown that a high antenna scan rate and a small number of hits per beam width are desirable for radars to be used with track-while-scan equipment.

It has also been shown that a slower antenna scan rate and more hits per beam width are desirable for systems with signal integration. Greater target resolution is possible with signal integration.

This paper has also presented a method for determining the width of the tracking gates which is based upon the mathematical expectation of tracking error.

The techniques presented in this paper represent another step in the direction of a safe, automatic Air Traffic Control system.

Acknowledgement

The author is indebted to L. M. Hollingsworth, who directed the effort spent on this study, and to many of the members of the Air Force Cambridge Research Center who assisted by discussion and criticism in the preparation of this paper.

Appendix I

Aircraft Turn Radius (r)

During a banked turn part of the lift force is used to overcome centrifugal force

$$L \sin \gamma = WV^2/gr \qquad (30)$$

where
- L = lift force
- γ = angle of bank
- W = aircraft weight
- V = aircraft velocity
- r = turn radins
- g = acceleration of gravity

For a true-banked turn there is no skidding or slipping of the aircraft and the bank angle is given by

$$\cos \gamma = W/L = 1/G \qquad (31)$$

or

$$\sin \gamma = 1 - \cos^2 \gamma = 1 - (W/L)^2. \qquad (32)$$

Solving Equation (30) for r, and using Equations (31) and (32),

$$r = \frac{V^2}{g\sqrt{G^2 - 1}} \qquad (4)$$

Appendix II

Effect of Antenna Beam Pattern

It has been assumed in this report that the antenna gain was constant over the antenna beam width. This made it possible to define p, the probability that a given signal would exceed the threshold level, as a constant quantity for all n hits per beam width during any radar look at the target. If, within the beam width, the antenna gain varies as some function of the azimuth angle, the probability p will not be constant. Equation (12) for the solution of P_1 for systems without integration and Fig. 5 for systems with integration will not be applicable.

If the antenna gain factor A is a function of the azimuth angle α

$$A = A(\alpha) \quad ,$$

the idealized maximum radar range R_o will also be a function of the azimuth angle

$$R_o = R_o(\alpha) \quad .$$

In terms of the gain factor

$$R_o(\alpha) = R_{om} A(\alpha)/A_m \quad ,$$

where R_{om} is the maximum possible radar range and A_m, the maximum antenna gain factor.

For a given value of the false-alarm time, the probability p may be determined from Fig. 3 as a function of the azimuth angle

$$p = p(\alpha) \quad ,$$

corresponding to the value of $R/R_o(\alpha)$.

For systems without integration, the probability of detection during one scan P_1 will be given by

$$P_1 = 1 - \prod_{k=1}^{n} (1 - p_k) \quad ,$$

where $p_k = p_k(\alpha)$ is the probability that

the k_{th} signal of the n hits per beamwidth exceeds the threshold level.

The procedure for systems with integration is more complicated[4,5].

Appendix III

Evaluation of Infinite Series $\sum_{k=1}^{\infty} P_k \mathcal{E}_k$ for Circular Turns

From Equation (17)

$$\sum_{k=1}^{\infty} P_k \mathcal{E}_k = \mathcal{E}_1 P_1 \sum_{k=1}^{\infty} k^2 (1-P_1)^{k-1} \quad .$$

Consider the infinite series

$$y = \sum_{k=1}^{\infty} k^2 (1-P_1)^{k-1} \quad .$$

Letting

$$x = (1-P_1) \quad ,$$

$$y = \sum_{k=1}^{\infty} k^2 x^{k-1}$$

$$= 1+4x+9x^2+16x^3+\ldots$$

Integrating,

$$\int ydx = a_o+x+2x^2+3x^3+4x^4+\ldots,$$

where a_o is a constant of integration.

Transposing a_o, and dividing by x,

$$\frac{1}{x} \int ydx - \frac{a_o}{x} = 1+2x+3x^2+\ldots$$

Integrating again,

$$\int \frac{1}{x} (\int ydx)dx - \int \frac{a_o}{x} dx = a_1+1+x+x^2+x^3+\ldots$$

where (a_1+1) is a constant of integration.

The infinite series

$$1+x+x^2+x^3+\ldots.$$

is a geometric series with sum equal to $1/(1-x)$.

$$\int \frac{1}{x} (\int ydx)dx - \int \frac{a_o}{x} dx = a_1+1/(1-x) \quad .$$

Differentiating,

$$\frac{1}{x} \int ydx - \frac{a_o}{x} = \frac{1}{(1-x)^2} \quad .$$

Multiplying by x,

$$\int ydx = a_o + \frac{x}{(1-x)^2} \quad .$$

Differentiating,

$$y = \frac{1+x}{(1-x)^3} = \frac{2-P_1}{P_1^3} \quad .$$

Therefore,

$$\sum_{k=1}^{\infty} P_k \mathcal{E}_k = \frac{(2-P_1)\mathcal{E}_1}{P_1^2} \quad , \qquad (19)$$

for aircraft targets performing circular turns.

Appendix IV

Evaluation of Infinite Series $\sum_{k=1}^{\infty} P_k \mathcal{E}_k$ for Sudden Course Changes

From Equation (22)

$$\sum_{k=1}^{\infty} P_k \mathcal{E}_k = \mathcal{E}_1 P_1 \sum_{k=1}^{\infty} k(1-P_1)^{k-1} \quad .$$

Consider the infinite series

$$y = \sum_{k=1}^{\infty} k(1-P_1)^{k-1} \quad .$$

Letting

$$x = (1-P_1) \quad ,$$

$$y = \sum_{k=1}^{\infty} kx^{k-1}$$

$$= 1+2x+3x^2+4x^3+\ldots$$

Integrating,

$$\int ydx = a+x+x^2+x^3+x^4+\ldots,$$

where a is a constant of integration.

Therefore,

$$\int ydx - a+1 = \frac{1}{1-x} \quad .$$

Differentiating,

$$y = \frac{1}{(1-x)^2} = \frac{1}{P_1^2} \quad .$$

Therefore,

$$\sum_{k=1}^{\infty} P_k \mathcal{E}_k = \frac{\mathcal{E}_1}{P_1} \qquad (23)$$

for an aircraft target performing a maneuver which involves a sudden change in course.

References

1. J. S. Hall, "Radar Aids to Navigation," Radiation Laboratory Series, Vol. 2, pp. 299-311, McGraw-Hill, 1947

2. B. F. Green, Jr., "Volscan," (Air Traffic Control Central AN/GSN-3), Air Force Cambridge Research Center, Cambridge, Mass., Oct. 1953

3. B. M. Jones "Elements of Practical Aerodynamics," pp. 364-377, Wiley, New York, 1950

4. J. I. Marcum, "A Statistical Theory of Target Detection by Pulsed Radar," Report No. RM-753, Rand Corp., Santa Monica, Calif., 1948

5. J. I. Marcum, "A Statistical Theory of Target Detection by Pulsed Radar-Mathematical Appendix," Report No. RM-754, Rand Corp., Santa Monica, Calif., 1948

6. S. M. Kaplan and R. W. McFall, "Statistical Properties of Noise Applied to Radar Performance," Proc. IRE, Vol. 39, pp. 56-60, Jan. 1951

7. J. V. Harrington and T. F. Rogers, "Signal-To-Noise Ratio Improvement Through Integration in a Storage Tube," Proc. IRE, Vol. 38, pp. 1197-1203, Oct. 1950

8. T. C. Fry, "Probability and Its Engineering Uses," Van Nostrand, New York, 1928

9. M. Ernst, "Inherent Automatic Tracking Errors," Report No. 4-35, Air Materiel Command, Cambridge Field Station, Cambridge, Mass., Dec. 1947

10. E. F. Grant, "A Study of the Prediction of the Position of a Flying Aircraft," Report No. E5042, Cambridge Field Station, Air Materiel Command, Cambridge, Mass., March 1949

11. L. N. Ridenour (ed.), "Radar System Engineering," Mass. Inst. Tech., Radiation Laboratory Series, McGraw-Hill, New York, 1947

12. J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," Mass. Inst. Tech. Radiation Laboratory Series, McGraw-Hill, New York, 1950

13. R. C. Spencer, "On the Errors in Linear Interpolation in R and θ Coordinates," Internal Memorandum, Air Force Cambridge Research Center, Cambridge, Mass., Aug. 1953

14. S. J. O'Neil, "Analysis of Some Pulsed-Radar Track-While-Scan Systems," Report No. E5080, Air Force Cambridge Research Center, Cambridge, Mass., Nov. 1951

15. L. V. Blake, "The Effective Number of Pulses Per Beamwidth for a Scanning Radar," Proc. IRE, Vol. 41, pp. 770-774, June 1953

A STUDY OF THE UHF OMNIDIRECTIONAL AIRCRAFT ANTENNA PROBLEM AND PROPOSED METHODS OF SOLUTION ≠

W. Spanos and J. J. Nail*
Federal Telecommunication Laboratories
a division of the International
Telephone and Telegraph Corporation
Nutley, New Jersey

The use of UHF (1000-3000 MC) aerial navigation equipment on large aircraft has resulted in antenna radiation problems due to the large size of the aircraft in wavelengths at these frequencies. Since the airplane is large relative to the wavelength, there is much shadowing of the radiation from antennas. Various parts of the airplane are located many wavelengths from the antennas and produce reflections of narrow lobe width. These factors tend to reduce the omnidirectional coverage in the 1000-3000 MC frequency band from single antennas on large aircraft.

By the use of electromagnetic modeling, antennas operating in the 1000-to 3000 MC frequency range were investigated at various sites on different types of aircraft[1]. It was found that a single antenna at any location on large aircraft would not give the required omnidirectional coverage in azimuth and + 30° coverage in elevation. Free space dipole range 15% to 50% of the time is the best omnidirectional coverage which may be expected depending on the location of the antenna.

Study of Dual Antenna Systems

Dual antenna systems were considered as a solution to the coverage problem. Dual antenna systems in which each antenna covers a 180° sector in azimuth were investigated on a DC-3. The wing tips and the nose and tail are the two sites which have a clear field around the antenna. The nose and tail antenna system was chosen as the best one to use since communication is usually to the front and rear of the airplane. Also there is more room to mount the antennas in the nose and tail. Patterns taken for a 1000 MC nose and tail antenna system are shown in Figure 1. For these patterns the nose antenna was deliberately given more gain than the rear antenna. The interference produced in the overlap regions between the two antennas is due to the large separation of the antennas, and for 1000 MC antennas on a DC-3, the lobes in this interference region repeat at about 0.8 degree intervals. Wingtip antennas have a similar pattern except that solid coverage is obtained off the wingtips and the interference region appears off the nose and tail.

Solution for DME

The Distance Measuring Equipment is not so stringent in its requirements upon an antenna system as the Radar Safety Beacon in several respects. The DME contains a memory circuit which keeps the equipment tracking during momentary losses of signal. Also the operation of the equipment does not necessarily have to be on an instantaneous basis. A top of the fuselage and a belly antenna could be operated one at a time through a switch[2]. The switch could be integrated with the equipment so that when the equipment fails to operate on the connected antenna, it switches to the other antenna. The preferred antenna would be the belly antenna since communication is usually from below the airplane. Both the top and the bottom antenna sites suffer from shadowing, therefore, with such a system, operating at 1000 MC on a DC-3, dipole range could be expected only about 70% of the time.

Nose and tail antennas could also be used through a switch. Figure 2 shows how the coverage would be provided. Each antenna has solid vertical coverage, therefore coverage approaching 100% could be expected. Nose and tail antennas might also be fed in parallel for DME operation. Operation in the interference region can best be determined by actual flight tests with the operating equipment.

Solution for Radar Safety Beacon

The Radar Safety Beacon type of equipment must maintain communication in all sectors on an instantaneous basis. A nose and tail antenna system in which each antenna covers a 180° sector will give solid coverage to the front and rear of the airplane; however, in the interference region off the wingtips, operation is uncertain. Figure 3 shows how this interference appears about the airplane. A statistical study of operation in this interference region was made in which frequency, spacing of antennas, airplane speed, distance to ground station and rotation rate

≠This project was sponsored by the Air Navigation Development Board and was administered by the Bureau of Aeronautics.
*Now with Bell Telephone Laboratories, Whippany, New Jersey.

of ground radar antenna were considered. If the flight of the airplane is smooth, the way in which the interference lobes sweep the ground station may be considered periodic. Since the rotation of the ground antenna is periodic the probability becomes that of periodic sampling of a periodic wave. Figure 4 shows the method used to compute probability. These probability figures are for 2700 MC antennas on a DC-6. P_1 is the probability of one hit in one try and $P(t)$ is the probability of a hit on the second try if the first try were a hit. $P_1P(t)$ is the probability of two successive hits. Figure 5 shows the probability of communication as a function of azimuth angle when the threshold level is 22 per cent of maximum range. This indicates that a nose and tail antenna system in which both antennas are fed in parallel is usable for Radar Safety Beacon operation if the lobe structure in the interference region is narrow enough. When this condition is met, then periodic phase shifting or switching of the antennas will not improve the performance.

Another solution for Radar Safety Beacon would be to feed top and bottom antennas in parallel with a low frequency phase shifter in one of the antenna feed lines. The lobe structure produced by a top and bottom antenna system is in the vertical plane with the lobes occurring at about 5° intervals. Figure 6 shows how this interference region occurs about the airplane. The worst interference occurs at the horizontal plane and without a phase shifter it would be possible for the airplane to fly such that minimum signal would be received at the ground station for a relatively long time. With a phase shifter in the system, the lobes scan the ground station and operation would then depend on probability which is determined by the spacing of antennas, frequency, distance to ground station, frequency of phase shifter, and rotation rate of the ground radar antenna. The total coverage provided would depend on the individual coverage figure for each antenna and the probability involved in the interference region. One of the major disadvantages of this system is that in the horizontal plane, which is the region of most importance, the probability is the worst and does not change appreciably with azimuth angle. Also the added weight and complexity of the phase shifter may not be desirable.

Common Systems for DME & Radar Safety Beacon

The use of two antennas in parallel such as in the nose and tail system lends itself well to the common operation of two equipments. Since the antennas present two loads to the equipment, a simple hybrid bridge type of multiplexer may be used. This allows the two equipments to feed two antennas without interference. Figure 7 shows how such a system would be connected.

Experimental Nose and Tail Antenna System

An experimental 1000 MC nose and tail antenna system was built for a DC-3. Figure 8 is a photograph of the nose antenna and Figure 9 is a photograph of the tail antenna used in the system. The nose antenna mounts inside a radome which replaces the nose section of the airplane. The tail antenna replaces the tail section of the airplane, the dipoles being protected by a radome cover. Each antenna is vertically polarized and consists of two stacked dipoles in front of a reversed 30° corner reflector. This arrangement provides 180° coverage in the horizontal plane and + 15° coverage in the vertical plane for each antenna. Figure 10 shows the superimposed free-space horizontal patterns of each antenna together with the region where interference occurs.

This antenna system was installed in the CAA DC-3 at the Technical Development and Evaluation Center in Indianapolis for the purpose of flight evaluation with the Distance Measuring Equipment and the Radar Safety Beacon equipment. Low loss 1/2" styroflex cable with a loss of 1.5 db was used to feed the two antennas in parallel.

Flight Tests for Nose and Tail Antenna System

Flight tests were made to determine the performance of the Distance Measuring Equipment with this antenna system. The ground equipment for this test consisted of a DTB transponder with a receiver sensitivity of about -90 dbm and a transmitter power of about 5 kw. The ground antenna was a four-element vertically polarized array with a gain of about 6 db. The antenna was located about 16 feet above the ground and the ground station was located about 800 feet above sea level. The airborne DME interrogator was a DIA previously modified to have a memory of about 20 seconds. The receiver sensitivity of the DIA was about -85 dbm and the transmitter power was about 1 kw. The speed of the airplane was about 150 mph for all flights.

Several types of flights were made in order to check the antenna system under various conditions. Closed circles (standard rate turns) of 1.5 miles diameter with 15-20 degree banks were flown to determine the omnidirectional coverage about the airplane in the + 15 degree vertical sector. Tangential paths to the ground station were also flown to determine the operation of the DME under the worst condition of flight when the interference lobes sweep past the ground station at

the slowest rate. The tests were extended to a range of 91 nautical miles (105 statute miles) in order to check the sensitivity of the system. Also at the longer ranges the interference lobes sweep past the ground station at an even slower rate which increases the duration of dropouts and decreases the performance of the system.

Tests were made using standard rate turns at various altitudes and ranges. The nose and tail system performed better than a 1/2 wavelength stub on the belly at all altitudes and ranges tested. At an altitude of 8000 feet above sea level and a distance of 91 nautical miles, the DME had momentary dropouts with the nose and tail system but operated continuously. The DME would not lock on the station when operated with the belly antenna. With a 1/2 wavelength stub on top of the fuselage, the DME had inadequate signal for a total (not consecutive) time of 30 seconds but it did not unlock. This test represented the worst condition for the nose and tail system for this type of flight since the range was large and the interference lobes swept past the ground station at a slower rate than at the closer ranges.

Tangential flights were made with the nose and tail system up to the maximum range of 80 nautical miles. During this flight the interference lobes off the wingtips in the region of $\phi = 90^{o}$ swept the ground station at the slowest rate. The long range test at 80 nautical miles represents the worst condition for the tangential flight since the interference lobes sweep past the ground station at the slowest rate. This means that the duration of signal dropouts in the minima is the greatest and the performance of the DME is the worst. A tangential path was flown, which was 30 nautical miles long and intersected the radial from the ground station at right angles. The flight was made at an altitude of 8000 feet above sea level and a range of 80 nautical miles. The DME operated continuously during this flight although there were momentary dropouts of signal every 5 to 6 seconds. Figure 11 is a portion of a recording taken during this flight. The recording was made by using the AGC voltage out of the DME. Dropouts are indicated on the recording by the minima regions and due to slow response of the recording equipment the amplitude does not drop to the threshold level. It is interesting to note the frequency of dropouts. Calculations show that the nulls sweep through the ground station at the rate of one per 30 seconds. Since the interrogate and reply frequencies are different their lobe structures differ and the number of dropouts should increase to 1 every 15 seconds. The recording indicates dropouts on the average of one every 5 to 6 seconds. The increased number of dropouts is probably due to the random motion of the airplane causing the lobes in the interference region to scan the ground station. This random motion of the airplane tends to increase the performance of this type of antenna system, since the rate at which the nulls appear at the ground stations increases and the duration of time in the nulls decreases.

A radial flight away from the ground station was made in order to check the gain of the nose and tail system relative to the stub on the belly. At an altitude of 1800 feet above sea level, the belly antenna operated the DME to a distance of 34 miles and the nose and tail system operated the DME to a distance of 41 miles. From this test and other observations it appears that the gain in the forward and rear sectors is about 2-3 db greater than a belly antenna.

A check of a 960-1215 MC hybrid bridge type of multiplexer was made with the nose and tail antenna system by operating two DME equipments simultaneously on the same DME ground station. This represents one of the worst conditions for multiplex operation. At an altitude of 8000 feet above sea level and a range of 85 nautical miles there was no decrease in performance for the two equipments when feeding the antenna system simultaneously.

Flight tests of this antenna system with the Radar Safety Beacon equipment are scheduled for the near future.

Comments and Conclusions

A study was made to determine the coverage provided by single and dual UHF antennas on large aircraft. As a result of this study, a nose and tail antenna system has been proposed as a solution to the UHF omnidirectional aircraft antenna problem. An experimental nose and tail antenna system has been built and has been tested for DME operation. Tests indicate that the nose and tail antenna system is satisfactory for DME operation. Any conclusions regarding Radar Safety Beacon must await further flight testing of the antenna system.

Acknowledgements

The writers wish to express their thanks to A. G. Kandoian and W. Sichak for the help and encouragement given during their supervision of this project.

References

1. W. Sichak and J. J. Nail, "UHF Omnidirectional Antenna Systems for Large Aircraft", Transactions of the IRE, Professional Group on Antennas and Propagation, January, 1954.
2. J. Hoffman, R. Carlson, W. Haworth, "Development of a DME Antenna Transfer Switch" CAA Technical Development and Evaluation Center, Report No. 201, May, 1953.

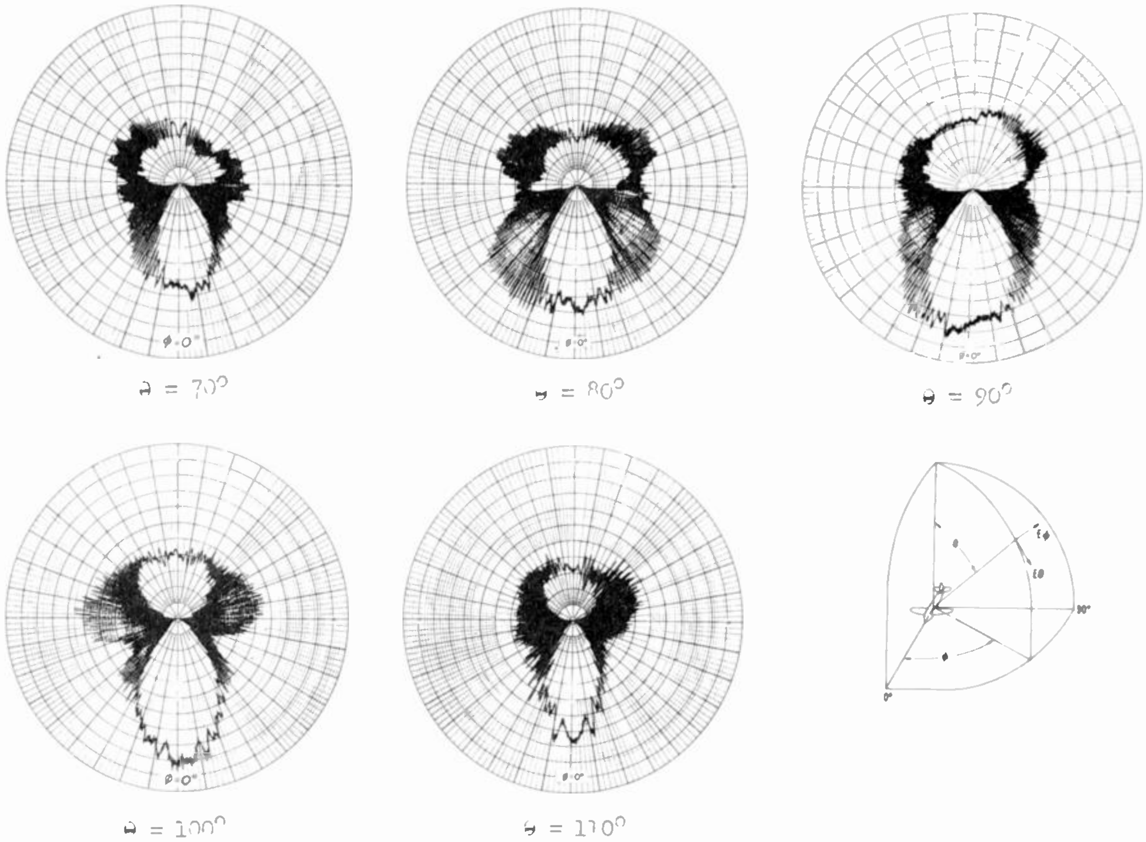


Fig. 1
Voltage patterns of horizontal-plane radiation from a discone antenna and 150-degree corner reflector at the nose and a discone and 300-degree reflector at the tail of a DC-3 aircraft. Measurements were made on a 1/24th-scale model at 24,000 megacycles to simulate behavior at 1000 megacycles on the actual aircraft. Wave polarization was in the E_θ plane. An aircraft superimposed on each polar diagram would have its nose pointed down.

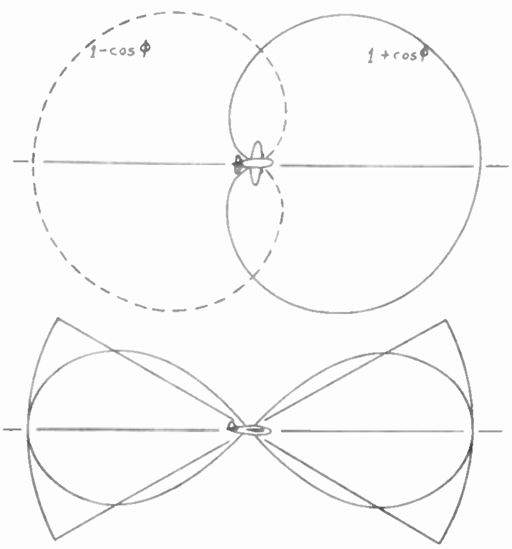


Fig. 2
Nose and tail sector antennas give the above cardioid patterns and may be switched to the equipment automatically to provide complete coverage. Their vertical patterns are given below, the half-power points occurring 30 degrees below and above the horizontal plane.

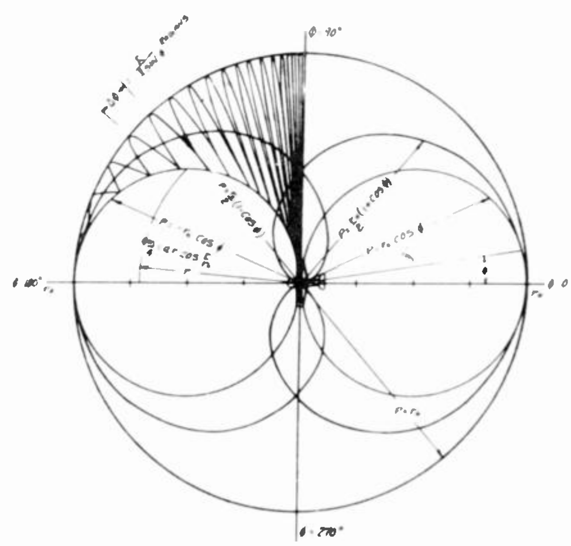


Fig. 3
Parallel operation of complementary cardioids. P = range, λ = wavelength, r_0 = maximum range, r = minimum usable range, l = distance between antennas, and $\varnothing_m$ = total angle where the range $\geq$ r.

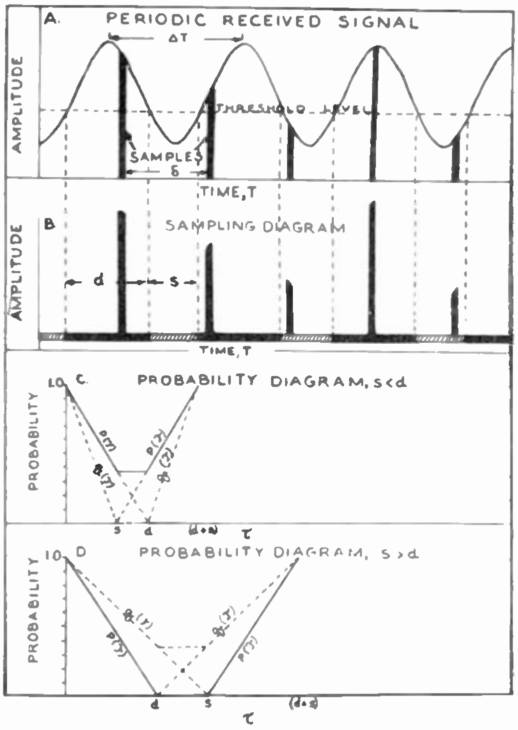


Fig. 4
Computation of probability of communication.

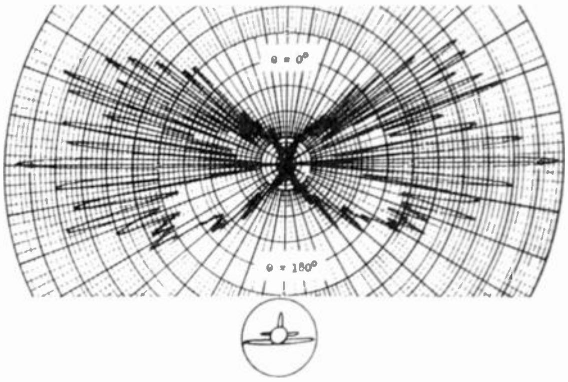


Fig. 6
Voltage pattern of vertical-plane radiation from two quarter-wavelength stub antennas mounted above and beneath the fuselage of a DC-3 aircraft.

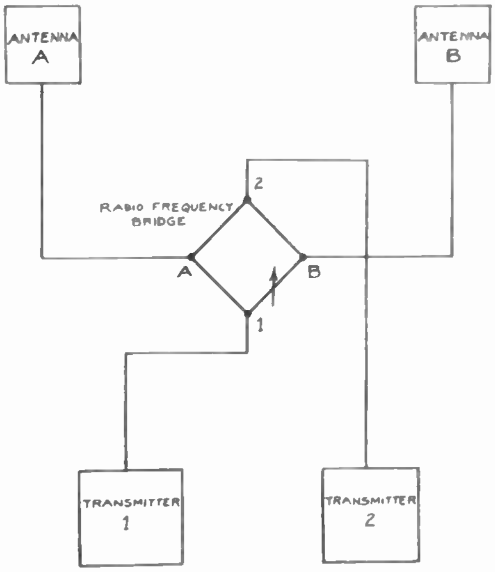


Fig. 7
Method of operating two equipments on a dual complementary parallel antenna system.

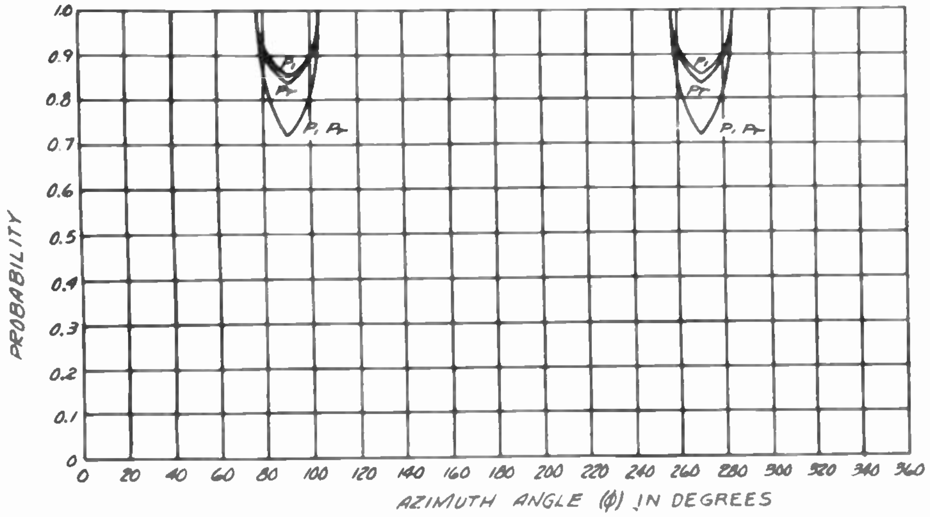


Fig. 5
Probability plotted against azimuth angle in degrees for the case where the usable threshold level is 22 per cent of the maximum range.



Fig. 8
Nose antenna for 960-1215-megacycle range.

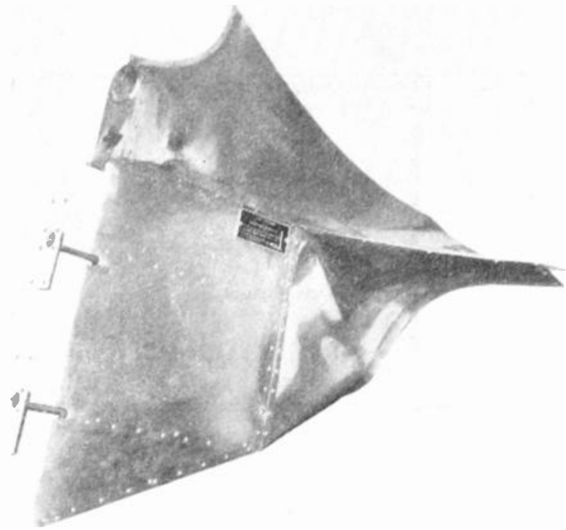


Fig. 9
Tail antenna for 960-1215-megacycle range.

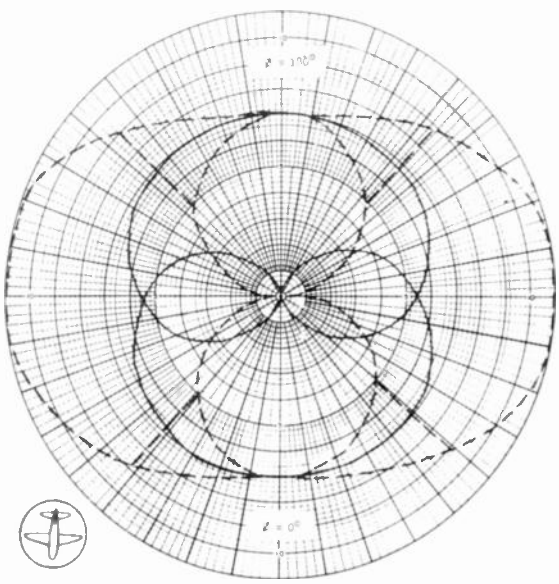


Fig. 10
Superimposed free-space horizontal radiation patterns of the nose and tail antennas are shown by solid lines. The outer dashed line indicates the region of maximum interference and the figure-of-eight inner dashed pattern is for minimum interference between the radiations of the two antennas. The radial dashed lines enclose the side regions of bad interference.

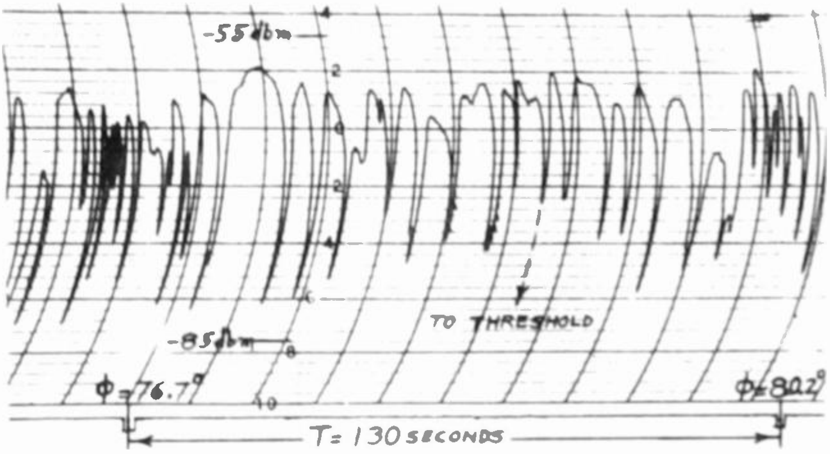


Fig. 11
Recording of automatic-gain-control voltage on a flight 80 nautical miles from and tangential to the distance-measuring-system transmitter.

A MODULATOR TECHNIQUE FOR PRODUCING SHORT PULSES IN HIGH POWERED MAGNETRONS

Thomas J. Parker
U. S. Navy Electronics Laboratory
San Diego 52, California

ABSTRACT

A "pedestal" technique is described for producing very short pulses (less than 0.1 microsecond duration) in high powered magnetrons. The pedestal technique applies a low-powered pulse, well rounded, to the magnetron as a pre-operating pulse bringing the magnetron r-f output to 5% to 10% of the full power level. Thereupon, a short high-powered pulse is applied to carry the magnetron output to the full power level.

Experimental circuits were constructed for pulsing the type 4J50 magnetron at pulse lengths of 0.15, 0.05 and 0.03 microseconds duration. Magnetron current, r-f envelope and r-f spectrum oscillograms were recorded for each pulse duration.

SUMMARY

The problem of producing in high powered magnetron applications short pulses of duration less than 0.1 microsecond has in the past generally been resolved by use of hard tube modulator techniques which require direct voltages in excess of the magnetron operating voltages.

In the line-type modulator, the pulse developed by the two-way time of travel of a voltage wave in a pulse forming network is coupled to the magnetron load by means of a voltage step-up transformer. For the very short pulse case, however, the pulse applied to the magnetron is established by the frequency pass-band characteristic of the pulse transformer while the pulse forming network degenerates into a single storage capacitor discharged by the thyratron into the transformer primary.

High powered magnetron operation presents a special problem in the formation of short pulses. For 4J50 magnetron applications using a line-type pulser, the voltage rate-of-rise rating limits the shortness of the pulse to be produced since a pulser with a single section network will produce a minimum duration pulse of about 0.3 microseconds. This value can be reduced as low as 0.2 microseconds by adding an integrating capacitor as well as a "tail biter" inductance across the input to the pulse transformer.

The pedestal technique applies a low powered pulse, well rounded, to the magnetron as a pre-operating pulse bringing the magnetron r-f output to 5% to 10% of the full power level. After the magnetron is oscillating due to the pre-operating pulse, a short high powered pulse is applied to carry the magnetron output to the full power level. Since the magnetron is already in the oscillating mode at the time of application of the high powered pulse, the rate of rise of magnetron current and r-f output power will be limited almost entirely by the rate of build up of the rotating electron cloud and by the "Q" of the resonant cavities and output circuit of the magnetron, this build-up time being in the order of a few millimicroseconds.

Experimental pedestal modulators were constructed for pulsing the type 4J50 X-band magnetron at pulse lengths of 0.15, 0.10, 0.05 and 0.03 microseconds duration. These experimental modulators in reality consisted of two modulators. Modulator No. 1 develops a pulse adjustable in amplitude and of 0.5 microseconds duration. A Darlington Line Modulator (See Glasoe and Lebacqz, Pulse Generators, MIT Radiation Laboratory Series Vol. 5, McGraw-Hill Book Company, Inc., 1948, Par. 464-465) produces the short high powered pulse achieving therein a voltage step-up without requiring the use of a pulse transformer. Timing pulses for these two modulators are established in a variable delay unit to cause the short pulse to arrive at the magnetron just after the start of oscillations due to the pulse from Modulator No. 1.

Measurements to determine modulator performance include oscilloscope photographs of magnetron pulse current, r-f envelope and r-f spectrum waveforms at each pulse duration.

Two improved one-modulator circuit arrangements, one for a 0.1 microsecond pulse and the other for a 0.05 microsecond pulse, have been designed and are under construction. These modulators will use single thyratron switch tubes and will establish all time sequencing in passive elements. It is expected that experimental data regarding the performance of these pulser circuits will be available prior to March 1954.

INTRODUCTION

The problem of producing in high powered magnetron applications short pulses of duration of less than 0.1 microsecond has in the past generally been resolved by the use of hard tube modulator techniques. These hard tube modulators have many advantages such as being capable of operating into magnetrons with a

wide range of impedance mismatch yet still forming good rectangular pulses. Also the pulse duration can be changed readily by changing low voltage components. The hard tube modulator also can produce series of short pulses with very close spacing.

The hard tube modulator, while in many respects quite flexible, requires a high "overhead"; meaning that the plate power supply voltage must exceed the peak voltage to be applied to the magnetron oscillator. Also, large transmitting type vacuum tubes must be used to pass the peak value of current drawn by the magnetron during its pulse. These two factors alone require that the hard tube modulator be rather bulky in space and inefficient due to plate power dissipated in the keyer tubes in passing large pulses of current.

The constant voltage characteristic which permits operation of the hard tube modulator into a wide range of magnetron impedances is, however, often responsible for reduced life and impared performance of the magnetron. This condition develops since occasional arc-overs which are inherent in high powered magnetron operation will cause very high cathode current to flow and thus damage the magnetron cathode with an attendant reduction in life.

Modulators of the hard tube type have been used successfully to produce pulses as short as .03 microseconds at power levels of approximately 200 KW r-f output power. In one such modulator[1], the keyer tubes were four 6D21 tetrodes in parallel driven by a keyer pulse formed by discharging energy stored in a short length of coaxial cable into a grid resistor by the switching action of triggering a 4C35 hydrogen thyratron.

The conventional line type pulser shown in Figure 1 has been used for the formation of modulator pulses also of the order of .03 microseconds; one such modulator pulsing the magnetron oscillator at approximately 30 KW output at K Band[2]. With the conventional line type pulser the pulse duration is established in the pulse forming network, this pulse duration being the round trip time of travel of a voltage wave impressed upon the network. In this function the pulse forming network can be either a real or an artificial transmission line. For quite short pulses with high rates of rise of the voltage pulse a real section of transmission line is often used, where, for a typical 50 ohm coaxial cable, the cable lengths needed are approximately 217 feet for a 1 microsecond pulse, 10 feet for an .05 microsecond pulse and 6 feet for an .03 microsecond pulse.

In the line type modulator of Figure 1 the pulse developed by the two way time of travel in the artificial transmission line serving as pulse forming network is coupled by means of a voltage step up transformer to the magnetron load. For the very short pulse case, however, the pulse applied to the magnetron is established by the frequency pass band characteristic of the pulse transformer matching the impedance of the pulse forming network to the impedance of the magnetron load and the pulse forming network degenerates into a single storage capacitor discharged by the thyratron into the transformer primary.

Consider the rectangular pulse of duration spaced at a repetition frequency, f_r. The frequency components present in such a pulse form the line spectrum of Figure 2 where the amplitude, A_K of any k-th harmonic of the pulse repetition frequency, f_r, is given by the equation

$$A_K = 2E_0 f_r \delta \frac{\sin \pi \delta k f_r}{\pi \delta k f_r}$$

The nulls observed in Figure 2 occur at frequencies where the A_k vanish, that is at the frequencies where kf_r equals $1/\delta$, $2/\delta$, n/δ, etc. For the restricted bandwidth case, such as presented by the pulse transformer, the output pulse shape deviates from rectangular. Figure 3 shows the comparative pulse shape when the actual pulse is restricted to contain frequency components below (a) $1/\delta$, (b) $2/\delta$ and (c) $3/\delta$. From this it is demonstrated that frequency components of the order of $2/\delta$ to $3/\delta$ should be passed by the pulse transformer. This means therefore that for a pulse of .1 microsecond that the transformer should pass at least frequencies up to 20 megacycles without loss or distortion if a reasonably near rectangular pulse is required.

This bandwidth requirement can often be met for medium or low power transformers with an upper limit of say 50 KW peak and for pulses of the order of .1 microsecond duration. For higher pulse powers, however, the design considerations for efficient power transfer mitigate against wide bandwidth. The larger core sizes and wire sizes and the heavier interlayer insulation cause an increased leakage reactance as well as increased dielectric losses both of which factors reduce the effective frequency pass band characteristics of the transformer. Also the problem of saturation in the outer laminations with magnetic flux barely penetrating to the innermost laminations restrict the upper power level as well as the upper pass band frequency. For these reasons the upper practical limit for pulse transformer operation at medium and high power has been for pulse durations greater than approximately .1 microsecond.

Consider next the case where the pulse forming network is allowed to discharge directly into the magnetron, this procedure eliminating the lossy pulse transformer. The The'venin equivalent circuit for this case is shown in Figure 4. For conditions of impedance match, which case will cause the transfer of energy stored in the pulse forming network to

the magnetron load in a single rectangular pulse, the voltage developed across the load will be E_N. This requires that the network be charged to a voltage twice as high as reouired for full power level in the magnetron. This means, in fact, that for a magnetron requiring 22 KV pulse the pulse forming network must be charged to 44KV.

The High Powered Magnetron

High powered magnetrons present a special problem in the formation of short pulses. Consider, for example, the 4J50 magnetron which requires that the rate of rise of voltage at the 80% point of the voltage pulse not exceed 110 KV per microsecond. Should the voltage pulse to this magnetron exceed this rate of rise, there is a marked tendency for the electron cloud not to form in the π-mode, this causing the dynamic impedance of the magnetron to be improperly high, which because of the modulator load characteristic also causes the magnetron pulse voltage to be improperly high.

The end result occasioned by a too rapidly applied voltage pulse to a high powered magnetron will be that of the magnetron failing to establish oscillation in its π-mode. With the line type pulser this causes the voltage to rise to a value considerably greater than normal, since the magnetron will either be operating as a diode or will be carried to a cutoff condition. Figure 5 illustrates the V-1 curve of a typical magnetron with the pulser load characteristic drawn through a typical operating point of the magnetron. From this figure where curve (a) represents the line type pulser load characteristic it can be seen that if the magnetron fails to operate in its proper mode and thus take power from the pulser that the pulse voltage applied to the magnetron will approach twice the normal operating value.

For the 4J50 magnetron application using a line type pulser the slope of the pulser voltage curve at about 80% of full voltage should not exceed the rate of rise of 110 KV per microsecond. This value of voltage rate of rise here limits the shortness of the pulse to be produced, since a pulser with a single section network will produce the minimum duration pulse of about .3 microseconds. This pulse width can be reduced slightly giving a pulse of .20 microseconds using the modified line type pulser of Figure 6. Here the energy is stored in the capacitor, C_1, and is discharged through the hydrogen thyratron, V_1, with the inductance L_1 being adjusted experimentally to establish the pulse duration. The capacitor C_2 serves to cause the voltage to the magnetron to be proportional to the integral of current through the thyratron thus allowing the rate of rise of voltage to be controlled during the pre-oscillating period. The inductor L_3 serves as a "tail biter", to cause the magnetron pulse to be terminated somewhat more sharply than otherwise would be the case. The inductor L_2 serves to isolate the capacitance in the filament circuit of the held off and back current diodes and thus reduce thyratron spike current.

The modulator circuit described in the preceding paragraph indicates that an approximate limit has been reached for short pulse high power modulators of the line type. The modulator technique to follow has been used at the U. S. Navy Electronics Laboratory to produce even shorter pulses at the full rated power for the particular high powered magnetron.

PEDESTAL VOLTAGE TECHNIQUE

The pedestal voltage technique requires that a low powered pulse, well rounded, be applied to the magnetron as a pre-operating pulse. This pulse brings the magnetron to the threshold of operation with a wave shape which allows the electron cloud to become sorted into the π-mode and to be producing RF power at about 5% to 10% of the full power level.

After the magnetron is passing current due to the pre-operating pulse, a short high-powered pulse is applied to carry the magnetron voltage to the full power level. Since the magnetron is already in the oscillating mode at the time of application of the high-powered pulse, the rate of rise of magnetron current and also of RF power will be limited almost entirely by the rate of build up of the rotating electron cloud and by the Q of the resonant cavities and of the output circuit of the magnetron, this time being of the order of a few millimicroseconds.

A simple circuit for the production of the short pulse on a pedestal is shown in Figure 7. The pedestal is formed by discharging the single section network upon firing Thyratron V1. For the circuit values shown this produces a magnetron current pulse approximately 1 microsecond in duration and approximately a half sinusoid in shape. At a time approximately .6 microsecond later Thyratron V2 is triggered to discharge the 50 ohm network #2 to produce a .15 microsecond pluse in the 4J50 magnetron.

Figure 8a shows the magnetron current waveform observed due to the pedestal pulse. This waveform is to the scale of 1 ampere per vertical division. The horizontal sweep scale reresents .4 microsecond per division. The magnetron current is carried to approximately the 1.6 ampere level at which value the jump in magnetron current indicates build up of oscillations in the π-mode. Figure 8b is the same as waveform Figure 8a but with the vertical scale set to 11.5 amperes per division. Figure 8c shows the magnetron current developed by triggering Thyratron V2 at approximately .6 microsecond after the triggering of

Thyratron V1. This second triggering action allows the network #2 to discharge into the magnetron causing its current to increase to the operating power level of approximately 30 amperes.

The RF pulse envelope produced by the .15 microsecond 30 ampere magnetron current pulse is shown in Figure 9a where the sweep scale is .1 microsecond per division. For comparison Figure 9b shows the RF power produced by the pedestal alone. Figure 10 shows the RF spectrum as viewed with a TS-148 spectrum analyzer. The frequency scale shows 12 megacycles between principal nulls.

The pedestal modulator of Figure 7 presents three serious drawbacks; namely, (1) two thyratrons are required, (2) accurate trigger sequencing is required and (3) the lower limit of short pulse operation is established by the pass band characteristics of the pulse transformer.

This latter effect, in fact, was demonstrated forcefully in the original test set-up wherein a pulse transformer designed for operation with a .25 microsecond pulse was used. For plate input voltage of 6.75 KV with resonance charging of the two section network #2, the full magnetron current of 30 amperes was observed. Reducing network #2 to a single section network by removing capacitor #2 shortened the pulse by approximately 50% as was anticipated but for no change in plate input voltage the magnetron current was observed to be only 15 amperes, a reduction in peak RF power of approximately 50%.

For pulse durations shorter than 0.15 microseconds duration, an alternate experimental modulator arrangement was necessary. This alternate circuit, shown in Figure 11, again consists of two modulators. Modulator No. 1, adapted from the Navy Model SU-1 radar, develops a pulse adjustable in amplitude and of 0.5 microseconds duration. A Darlington Line Modulator[4] produces the short high-powered pulse achieving therein a voltage step-up without requiring the use of a pulse transformer. Timing pulses for these two modulators are established in a variable trigger delay unit to cause the short pulse to arrive at the magnetron just after the start of oscillation due to the pulse from modulator No. 1.

A more complete discussion of the functioning of the Darlington Line modulator is presented in later paragraphs and in Appendix 1. Suffice it to say here that the Darlington Line modulator used in the pedestal modulator experiments was constructed of four five-section networks, each network being an artificial transmission line which has a two-way time of travel of 0.1 microsecond.

The waveforms observed for the 0.1 microsecond pedestal modulator are shown in Figure 12, this figure showing (a) magnetron current, (b) r-f envelope and (c) r-f spectrum.

For the experimental modulators to produce the 0.05 and the 0.03 microsecond pulses, the individual artificial network were reduced to two section networks and to one section networks respectively.

The variable trigger delay unit was in each case adjusted to cause the short pulse to arrive at the magnetron at the peak of the much longer pedestal pulse. Magnetron current, r-f envelope and r-f spectrum waveforms for the 0.05 microsecond pulses are presented in Figure 13, while waveforms for the 0.03 microsecond pulses are shown in Figure 14.

A work here is necessary concerning instrumentation for observing the above waveforms:

1. Pulse magnetron current was observed by means of a current viewing resistor inserted between the magnetron anode and ground, while average magnetron current (not recorded here) was measured in the pulse transformer ground return. The stray capacitance to ground shunting the current viewing resistance causes the magnetron current waveform to be distorted due to the integrating effect of the resistance-capacitance network. This distortion while hardly noticeable in the 0.1 microsecond pulse case renders almost meaningless the current waveforms for the 0.05 microsecond and the 0.03. microsecond pulse cases. (Magnetron pulse current in the pedestal modulator cannot be observed in the pulse transformer ground return since the short pulse currents do not flow in that circuit).

2. The r-f envelope waveform was produced in a crystal detector where the rectified voltage was developed in the capacitance of the crystal holder shunted by a 50 ohm carbon-disc resistor. A cathode follower probe designed for operation with the Tektronic Model 517 oscilloscope was used to reduce the capacitance shunting the disc resistor. In this manner the rise-time and fall-time in the envelope detector waveforms are below 0.01 microseconds being probably nearer 0.001 microseconds.

3. R-f spectrum waveforms were recorded from a Model TS-148 spectrum analyzer. While this spectrum analyzer is fully satisfactory for operation with pulses longer than about 0.1 microsecond, the usefulness of this spectrum analyzer is quite restricted for the very short pulse cases. First, the intermediate frequency of the TS-148 analyzer is 22 megacycles which allows image frequency components to distort the spectrum of the very short pulse. Secondly, the sweep frequency range is about 50 megacycles to give a further distortion of the very short pulse spectrum.

Note that in Figure 14 (c) the r-f envelope rise-time and fall-time each are of the order

of 0.01 microseconds which value is consistent with pulse shapes produced by single section pulse forming network. This suggests that the rate of build-up of r-f energy in the cavities and output circuit of the magnetron will be somewhat less than this value. If this be the case, it is a logical extrapolation to suggest that the minimum pulse duration obtainable will probably be of the order of 0.01 microseconds. (It is worthy of note here that the type 4J50 magnetrons used in these tests were manufactured by Litton Industries, San Carlos, California under the Jan. 1A Specification dated 30 January 1952).

The Darlington Line[4]

The Darlington Line to be described in succeeding paragraphs offers a technique which can be employed to overcome the three disadvantages listed previously for the pedestal modulator of Figure 7. The Darlington Line employs the characteristic that where a voltage wave traveling along a transmission line meets a discontinuty, the boundary conditions at the discontinuity cause the magnitude of the wave traveling down the line to be changed as well as to cause a voltage wave to be reflected back toward the source.

Transmission line theory gives the magnitude of the reflected wave as

$$V_r \frac{(R_L - Z_o)V}{(R_L + Z_o)}$$

where R_L is the load impedance and V is the voltage traveling down the transmission line of characteristic impedance, Z_o. The voltage developed across the load impedance is established considering the The'venin equivalent circuit. The voltage wave of V volts reaching an open circuit load will cause a voltage wave of 2V volts to appear at the open circuit. The The'venin equivalent circuit thus becomes that shown in Figure 15 and the voltage developed across the load is $V_L = \frac{2R_L V.}{Z_o + R_L}$

From the above relation it can be seen that there will be a voltage magnification where $R_L > Z_o$.

This voltage step feature is employed in the Darlington Line shown schematically in Figure 16 as being composed of n networks of artificial transmission line. Each network has the same two way time of travel, , of a voltage wave. The characteristic impedances of the various networks are chosen as follows:

$$Z_n = \frac{R_L}{n}$$

$$Z_k = \frac{k(k+1)}{n^2} R_L \quad k = 1, 2____, (n-1).$$

When initially all capacitors are charged to a voltage, V_n, the closing of the switch S_1 will produce a voltage pulse of duration δ and amplitude $\frac{nV_n}{2}$ across the load resistor R_L. The switch S_1, in this case may be either a spark gap, a gas discharge tube or may be a hydrogen thyratron switch tube.

The voltage waves developed at various times through the Darlington Line are shown in Appendix A, where it is established that the voltage pulse appearing at the load is delayed in time from the instant of firing of the thyratron. This time delay is established by the time of travel of the wave generated by the firing of the thyratron down the artificial transmission line to load. This delay time, D, is

$$D = \frac{(n-1)}{2}\delta \text{ microseconds}$$

where D is expressed in microseconds.

Pedestal modulator using Darlington Line

The delay produced in the Darlington Line can be used to establish the time interval between the pedestal and the short pulse. Consider the circuit of Figure 17 using a 4 network Darlington Line to produce a modulator pulse of .1 microsecond duration. Each individual network is a 5 section network and during pulsing combine to produce the rectangular pulse of Figure 18.

This pedestal modulator of Figure 17 is designed to pulse the 4J50 magnetron for which the rate of rise of pulser voltage must not exceed 110 KV per microsecond, and for which the voltage-current characteristic of Figure 19 (a) apply. The magnetron operating point, P, presents a static load of 800 ohm and a dynamic load of 235 ohm to the pulser. A crude equivalent circuit of this magnetron load, as far as pulser is concerned, is shown in Figure 19 (b).

The single section network in Figure 17 is chosen to produce a half sine wave of voltage for the magnetron while in the diode current region. This voltage wave will be

$$e_p = E_m \sin t\omega$$

where E_m represents the magnetron voltage at the threshold of operation in the π-mode, and ω is the radian frequency of oscillation of the single section L-C network discharging energy into the diode resistance of the magnetron. For the 4J50, E_m = 16 KV approximately, allowing ω to be chosen to meet the voltage rate of rise requirement of the magnetron. The rate of rise of the pedestal voltage is

$$\frac{de_p}{dt} = \omega E_m \cos \omega t$$

and must not exceed 110 KV per microsecond of the threshold voltage, 16KV. Choosing as a safe value the slope at the 70% voltage level we get

$$\frac{de_p}{dt} = 110X10^9 = \omega x\ 16000\ x.7$$

which neglecting the diode operation loading gives as a realistic value of frequency for the L-C network of 1.5 megacycles or a period of .33 microseconds for the pedestal voltage pulse.

The triggering of the hydrogen thyratron V_1 initiates the action of the single section network in producing a 16KV pedestal voltage which reaches its maximum value at a time .17 microseconds after the triggering of the thyratron. During the formation of the pedestal, the voltage wave produced by the firing of the thyratron travels down the Darlington Line and is transformed up in voltage. By the time the voltage wave reaches the end of the third network in the Darlington Line, it is delayed in time by three times the one way time of travel (or by three halves the pulse duration) which for a .1 microsecond pulse is at a time .15 microsecond after the triggering of the thyratron V1. The short pulse produced in the Darlington Line thus is delayed in time to arrive at the peak of the pedestal pulse and is coupled to the magnetron to produce a 250 KW pulse of .1 microsecond duration.

Note here that the firing of a single thyrathron produces both pedestal and power pulse and that the time sequencing is established totally by passive elements. Not, moreover, that the short pulse is developed without requiring a pulse transformer. The pedestal pulse, however, is coupled to the magnetron through a pulse transformer to permit the single section pedestal network to be charged in parallel with the Darlington Line. This transformer passes a .33 microsecond half-sinusoid pulse at a rather low power level, while serving, as well as the d-c return path for the magnetron cathode circuit.

CONCLUSION

Experimental investigations at the U. S. Navy Electronics Laboratory have verified the principle of the pedestal technique for pulsing the high powered magnetron where a voltage rate of rise specification restricts the shortness of pulse desired. While the investigation reported herein were restricted to the pulse durations of 0.15 microseconds, 0.10 microseconds, 0.05 microseconds and 0.03 microseconds, it is believed that the minimum pulse duration as established by the rate of build up of the rotating electron cloud and of r-f energy in the cavities and output circuit has not been reached.

This pedestal technique is suggested for applications with the older style magnetrons which give erratic performance making possible in this manner the utilization of magnetrons now being considered inoperable. The pedestal technique is also suggested as a device by which parallel operation of two or more magnetrons will become a practicality. Since all magnetrons in this case may be pre-pulsed simultaneously thus to establish a common operating frequency, it is felt that the common operating frequency will be maintained when the input voltage is raised to the full power level.

ACKNOWLEDGEMENT

The author wishes to express his appreciation for the suggestions and assistance of the co-workers in the Radar Equipment Section, U. S. Navy Electronics Laboratory and of Dr. Norman Moore and Mr. Paul Crapuchettes of Litton Industries leading to the development of the pedestal modulator technique.

APPENDIX A

THE DARLINGTON CIRCUIT

The block diagram of the Darlington circuit is shown in Figure A-1. It consists of (n-1) four terminal pulse forming networks while the n-th network is a two terminal pulse forming network. The requirements to obtain a single voltage pulse across the load are (1) that all the networks have the same delay time and phase characteristics and (2) the impedance relationships in the following pair of equations apply:

(1) $$Z_r = R_L \frac{r(r+1)}{n^2} \qquad r = 1, 2, \cdots (n-1).$$

(2) $$Z_n = \frac{R_L}{n}$$

For the energy stored in the network to be dissipated in a single pulse in the load, the reflections taking place between networks must of necessity cancel each other. The voltage of the pulse thus formed is equal to n/2 times the network voltage.

The voltage and current waveforms developed in the Darlington curcuit can be obtained from simple transmission line theory, as illustrated in Figure 2-A for a four network Darlington circuit. The action of closing the switch, S, produces a voltage wave equal to and of polarity opposite to the voltage E_n, initially across the network. This voltage wave e_1 travels down the first section unchanged until the discontinuity is reached between networks Z_1 and Z_2. Figure A-2(a) shows this voltage e_1 just before it reaches the junction between the two networks. Here two voltage waves are produced, one traveling

in the second network, the other reflected back in the first network as shown in Figure A-2(b). The The'venin equivalent circuit can be applied at the junction using the characteristic impedances of the two networks and noting that the open-circuit voltage at the junction terminals will be twice the magnitude of the wave traveling down the network. The voltage e_2, traveling in the forward direction in the second network is

$$e_2 = 2e_1 \frac{Z_2}{Z_1 + Z_2} = \frac{3e_1}{2}$$

which is measured from the voltage level, E, to which the network was originally charged. The voltage wave e_1' traveling in the negative direction in network #1 is referenced to the zero voltage level since the initial voltage wave leaves the PFN capacitors uncharged. (Energy remains in the network however due to currents flowing in the inductor). The magnitude of e_1' is

$$e_1' = 2e_1 \frac{Z_1}{Z_1 + Z_2} = \frac{e_1}{2}$$

These waves travel in their respective directions to produce further voltage waves and reflections as scheduled in Figure A-3 for a four network Darlington circuit.

The primary voltage wave traveling down the Darlington circuit is amplified by the impedance change between networks to produce a pulse $\frac{n}{2}$ times the voltage, E_n, to which the network was initially charged.

Comparison now can be made to voltage multiplication by means of the Darlington circuit as compared to the conventional single network line pulser employing a pulse transformer to give the voltage step up. Assuming momentarily the use of a loss-less pulse transformer where the secondary to primary turns ratio is an integer, i.e., 2, 3, 4 etc., energy considerations require that the impedance of the pulse network be matched by the transformer impedance ratio to that of the load giving the voltage on the network as $E_n = \frac{2}{n} \cdot e_L$ which equation is identical to that of the n-network Darlington Circuit. Energy storage in the n-network Darlington circuit must therefore be identical to the energy stored in the single network of the conventional pulser giving thus that the total capacitance in the n-network Darlington Circuit will equal the total capacitance in the network of the conventional line type pulser employing a n:1 transformer turns ratio. Voltage to these two networks, moreover, will be the same as will the plate supply currents for a given application.

The Darlington Circuit offers a notable advantage for high power and/or short pulse operation of a magnetron oscillator since the impedance transfer is achieved through loss-less networks rather than by means of a transformer which is relatively more lossy. There is also a distinct weight saving since the relatively heavy transformer is eliminated. (Note that since the total capacitance of the n-network Darlington Circuit will equal the total capacitance in the n:1 transformer-coupled network the weight of the two networks will not differ appreciably.)

BIBLIOGRAPHY

1. Glasoe and Lebacqz - Pulse Generators, Radiation Laboratory Series Vol V, McGraw-Hill Book Co., 1948 -- pp 160-165

2. Modulator Techniques for Precision Radar, Cambridge Field Station Report No. E5041

3. Collins, G. B., Microwave Magnetrons, Radiation Laboratory Series Vol VI, McGraw-Hill Book Co., 1948

4. Glasoe and Lebacqz - Pulse Generators, Radiation Laboratory Series Vol V, McGraw-Hill Book Co., 1948 -- pp 464-465

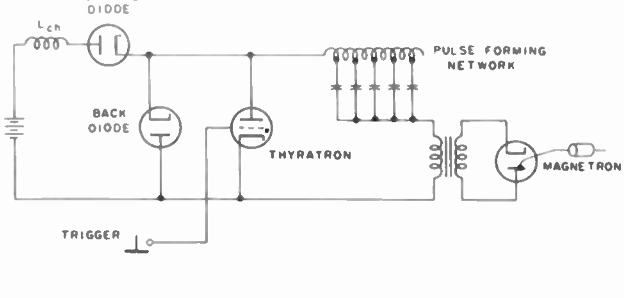


Fig. 1 - Line type pulser.

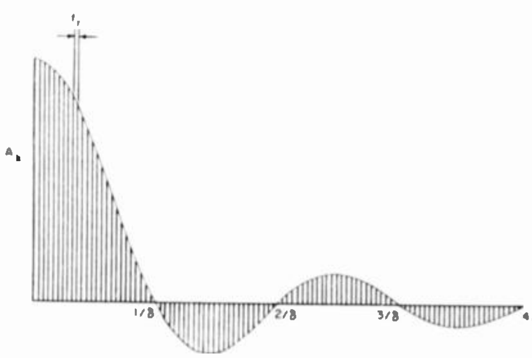


Fig. 2
Line spectrum for repetitive rectangular pulse.

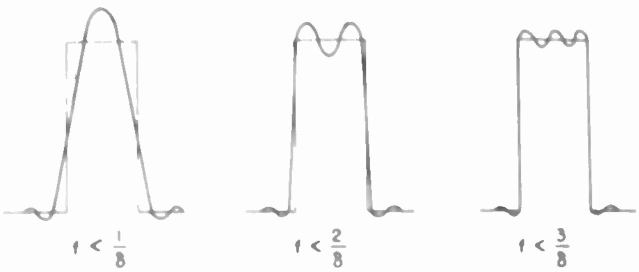


Fig. 3
Effect on rectangular pulses by restricting bandwidth.

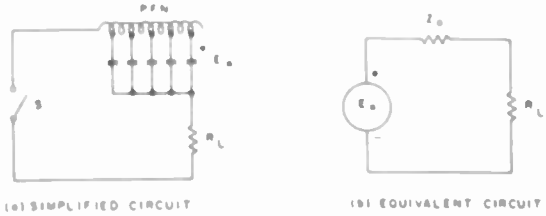


Fig. 4
Discharge circuit of pulse forming network.

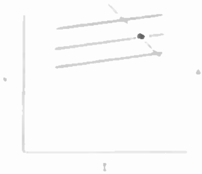


Fig. 5
Voltage-current diagram (idealized) for magnetron showing (a) pulser load characteristic.

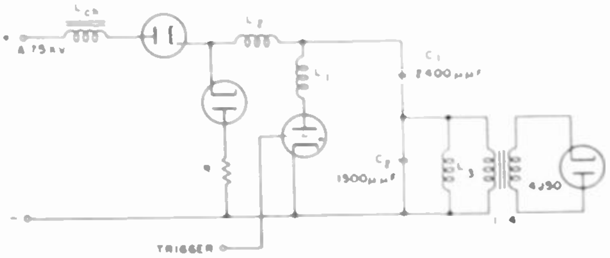


Fig. 6
Line type pulser for developing 0.25 μsec pulse with type 4J50 magnetron.

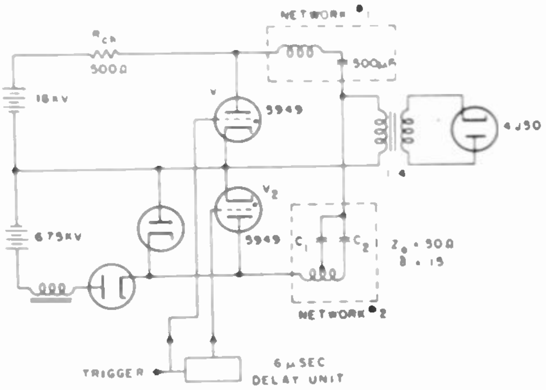
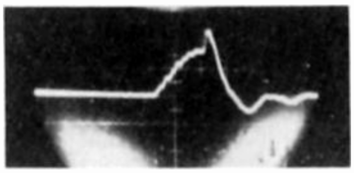
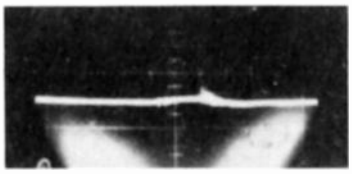


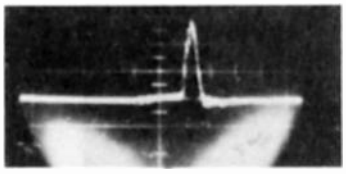
Fig. 7
Breadboard pedestal modulator.



(a) Pedestal only
1 amp/div

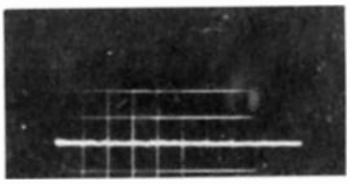


(b) Pedestal only
11.5 amp/div
Sweep speed - 0.4 μsec/div

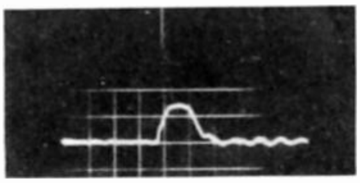


(c) Current pulse
11.5 amp/div

Fig. 8 - Magnetron current.



(a) During current pulse



(b) During pedestal only

Sweep speed - 0.1 usec/div

Fig. 9 - RF envelope.



(Sweep speed 12 mc between nulls)
Fig. 10
Spectrum for 0.15 μsec pulse with envelope shown in Fig. 9.

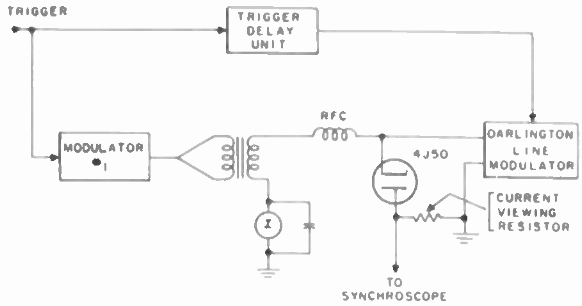
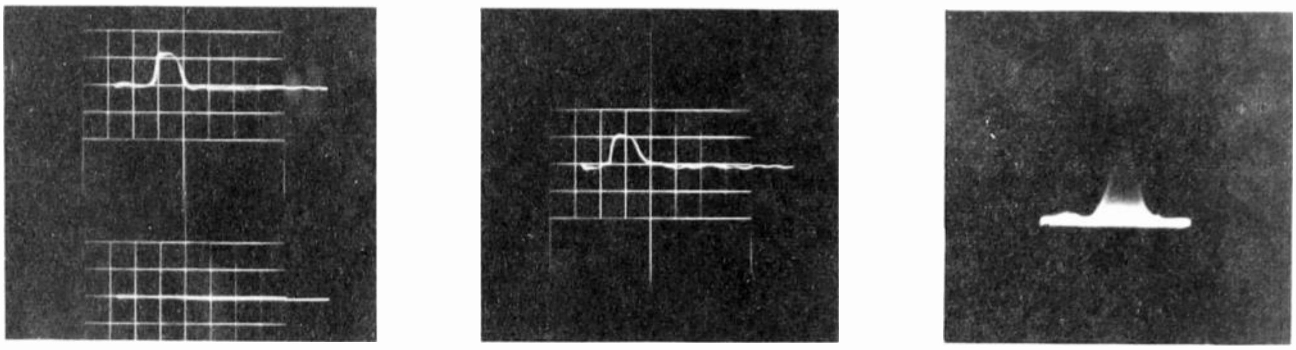


Fig. 11 - Block diagram of experimental pedestal modulator.



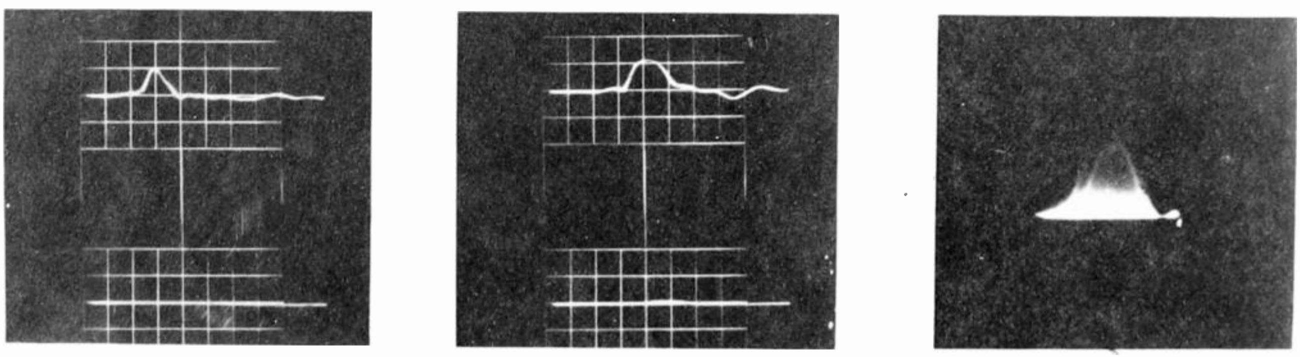
(a) Magnetron current

(b) RF envelope

(c) Spectrum
Approx 17 mc between nulls

Sweep speed - 0.1 μsec/div
(Lower trace for pedestal alone)

Fig. 12 - Waveforms for 0.1 usec pulser.



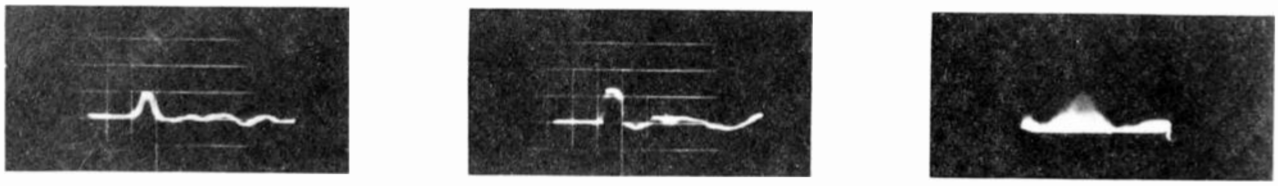
(a) Magnetron current

(b) RF envelope

(c) Spectrum
Approx 30 mc between nulls

Sweep speed - 0.05 μsec/div
(Lower trace for pedestal alone)

Fig. 13 - Waveforms for 0.05 μsec pulser.



(a) Magnetron current

(b) RF envelope

(c) Spectrum
Approx 45 mc between nulls

Sweep speed - 0.05 usec/div

Fig. 14 - Waveforms for 0.03 μsec pulser.

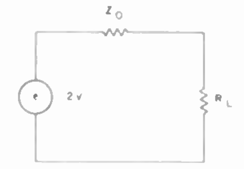


Fig. 15
Thevenin equivalent circuit of transmission line terminated in load resistance R_L.

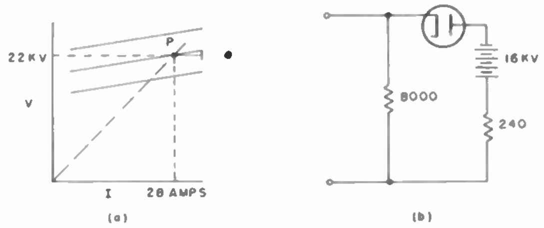


Fig. 19
(a) Magnetron characteristic (idealized); (b) magnetron equivalent circuit (values apply approximately for type 4J50 magnetron).

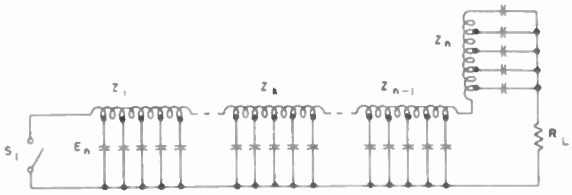


Fig. 16 - Darlington line pulser.

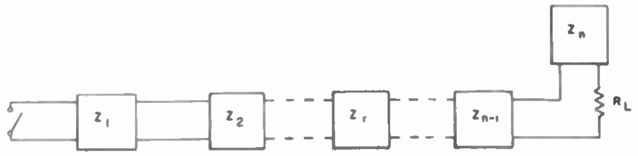


Fig. A-1
The Darlington circuit, block diagram.

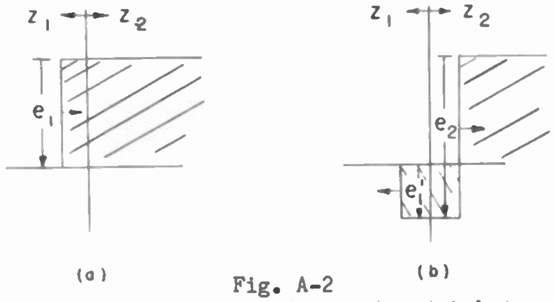


Fig. A-2
Voltage waves at impedance mismatch between Darlington circuit networks (a) voltage wave approaching junction and (b) voltage waves produced at impedance mismatch.

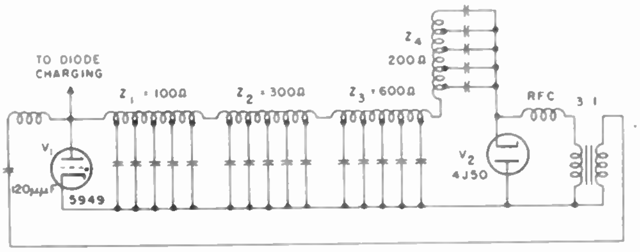


Fig. 17 - Pedestal - plus - Darlington line pulser.

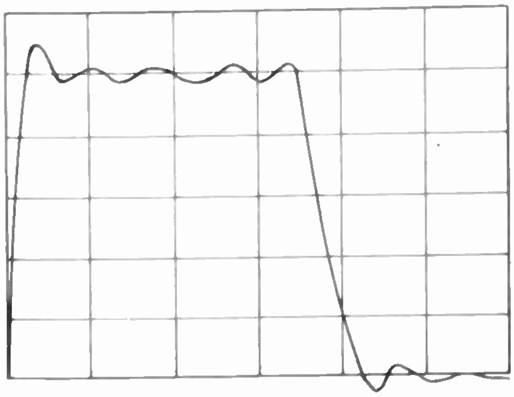


Fig. 18
Voltage pulse from 5-section pulse forming network.

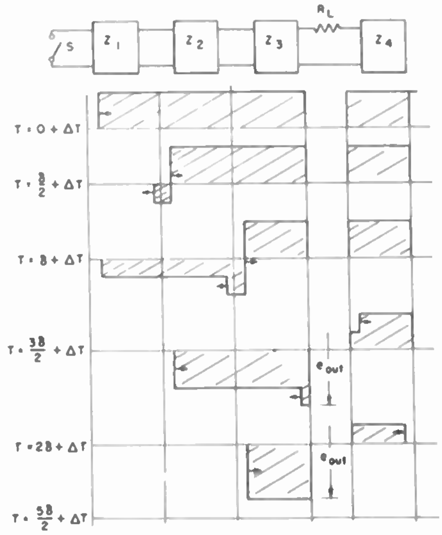


Fig. A-3
Voltage waveforms produced in pulsing a four-network Darlington circuit.

THE ROLE OF STEREO IN
"3-D" RADAR INDICATING SYSTEMS

By Walter R. Tower
Sperry Gyroscope Company
Division of The Sperry Corporation
Great Neck, New York

Summary

Indicator presentation of three-dimensional radar data has been accomplished for many years but, in many cases, in a less than ideal manner. A survey of generalized methods heretofore in use is given in this paper. Present day methods and contemplated systems are set forth. The stereo type of indicator, including the general theory of its operation and the human factors involved, is described. Examples of various optical arrangements are discussed. A typical stereo electronic circuit is described. Results of tests using an artificial indicator and an actual indicator operating with a search type radar are mentioned. Advantages and disadvantages of the system are brought out and application possibilities are stressed.

The Problem of
Three-Dimensional Presentation

Currently there is much popular interest in the problem of the presentation of three-dimensional information from a machine to a human observer. The motion picture industry has been expending considerable effort to capitalize on the three-dimensional appeal by such means as the peripheral vision screen (as used in "Cinerama" and "CinemaScope"), as well as those systems requiring the projection of separate eye images (the various "3-D" displays).

Basic Radar Displays

In past years various types of radar indicators have been developed to present to the observer spatial data derived from the radar antenna position and from the time in which echoes were received from reflecting objects. In many cases, however, the method of data presentation could be improved.

The basic types of indicators[1] which have been used are shown in Fig. 1. The A scan is essentially a one spatial dimension indicator presenting range horizontally on the tube face. The B scan presents two dimensions, range vertically and bearing (azimuth) horizontally. The C scan also presents two dimensions, elevation vertically and bearing horizontally. The familiar PPI presents range radially and presents bearing as the polar direction of a spot on the tube face. These indications are presented on the face of the cathode ray tube, which is, of course, a two dimensional surface. Therefore, the actual position of the radar data on the tube face can be specified at most in two dimensions, although three coordinates are required to describe completely the position of a reflecting object.

A two-dimensional indicator is adequate in many radar applications. For example, in the detection and location of naval vessels by other naval vessels, azimuth and range are sufficient to specify position. However, in cases such as the detection of aircraft by ground based radars, the two-dimensional display of the aircraft position leaves large areas of uncertainty as to the position of the aircraft.

Combined Displays

When positional information in three dimensions is desired for only a few reflecting objects, it is feasible to allow the operator to view separate two-dimensional indicators placed side by side. By correlating their information, he can deduce the three-dimensional space coordinates of the targets. Many combinations are possible in this method of presentation.

As an example of this method, Fig. 2 shows a type B and C combined indicator. The three-dimensional position of the target A can be deduced from a correlation of the two scopes as shown. In this type of presentation, it is often possible to increase the number of targets handled by various gating systems wherein only those targets which are manually gated on one scope are allowed to appear on the second scope.

Quantized Coordinate Display

In order to present the third dimension on a single two-dimensional surface it is sometimes possible to so quantize one dimension that the additional dimension can be shown within a region in which the quantized dimension is constant. Fig. 3a shows a type D indicator, in which coarse range information is provided by the position of a signal in the broad azimuth trace at various constant elevation levels. Such a presentation is applicable only when the

addition of the coarse third dimensional data (range in this case) justifies the loss of resolution in the quantized dimension (elevation in this case).

Coded Third Dimension

In many applications it is feasible to provide a rough type of third dimensional information by the use of special positional, numerical, or color codes at the position of the reflecting object on the CRT. Fig. 3b shows the positional code system known as a Type H indicator, which presents azimuth horizontally and range vertically. A radar target appears as two dots, the left dot showing range and azimuth, the adjacent right dot giving a rough indication of elevation.

The coding also may take the form of a numerical symbol as shown in Fig. 3c, which shows a type of indicator proposed for certain air surveillance radars. In this system a PPI which normally presents slant range data is modified to present a ground plane plot and the coded dots are used to signify the various height levels of the target to which they are adjacent.

Color also may be used as the coding means. The representation of each reflecting object may be depicted as a certain shade or color according to its spatial value in any one dimension.

Indicators in which one or more of the coordinate dimensions are coded may find considerable use when the positions of a large number of reflecting objects are to be presented.

Perspective Three-Dimensional Displays

Berkely[2] has shown that a single cathode ray tube may be used to present the illusion of three-dimensional space by means of the proper mixing of the three coordinate variables. For instance, suppose a cube is projected onto a two-dimensional surface. In an orthogonal view, it would appear as a two-dimensional figure. However, an oblique perspective* of the cube would appear as in Fig. 4. Then the three-dimensional variables (x, y, and z), of the cube are combined to give two-dimensional (horizontal and vertical) coordinates having the same origin and being geometrically defined by:

$$X = x - z \sin \theta$$

$$Y = y - z \cos \theta$$

where

X = horizontal coordinate in two-dimensional plane (CRT face)

Y = vertical coordinate in two-dimensional plane (CRT face)

x, y, z = coordinates in three-dimensional space

θ = angle the z-axis makes with the y-axis in projection

Physically, a simple representation of this transformation can be realized quite easily. For instance, suppose it is desired to represent the plane $y - c = o$ on a scope face. Choose x and z as the independent variables to cover the plane as two sweep voltages of different but integrally related frequencies. Then:

$$x = nkt$$

$$z = kt$$

where

t - time

k = slope of sawtooth

n = slope multiplier equivalent to frequency multiplier

Using the transformation equations with $\theta = 45°$:

$$X = x - z \sin \theta$$

$$\text{or } X = nkt - kt \sin 45°$$

and

$$Y = y - z \cos \theta$$

$$\text{or } Y = c - kt \cos 45°$$

To satisfy the above equations X (horizontal deflection voltage) can be a 3 kc sawtooth added to a 15 cps sawtooth while Y (vertical deflection voltage) can be a constant (positional) voltage plus the same 15 cps sawtooth. The plane $y - c = o$ then appears on the scope as seen in Fig. 5a.

By similar methods the three-dimensional surface $A \sin z - y = 0$ shown in Fig. 5b is produced by deflection voltages:

X = 1.5 kc sawtooth + 30 cps sawtooth

Y = 30 cps sawtooth - 60 cps sine wave

This type of three-dimensional presentation is quite effective where a limited volume of space is to be shown,

- - - - - - -

*The term "perspective" as used here is not strictly accurate, since, in a true perspective drawing, lines which are parallel in the figure to be drawn do not appear parallel in the perspective view but meet at some remote point.

and it has been used in several displays where radar antenna tracking errors were to be presented. One form of this type of tracking scope is shown in Fig. 6. The elevation error between the reflecting object and the radar antenna is shown as a vertical deflection plate voltage, the azimuth error is shown as a horizontal deflection plate voltage, and the range error is applied as equal voltages to both the horizontal and vertical plates. The three error voltages (azimuth, elevation, and range) then are effective in moving the spot away from the zero or no error position.

It is necessary to add various three-dimensional "clues" to a representation of this type in order to increase the depth effect and to avoid interpretation ambiguities. In Fig. 6 a three coordinate axis has been added and the X and Y deflection path of the target has been made visible to render a single interpretation of the position in space of the target relative to the zero error position.

This type of indication is adequate where there are only a few reflecting objects to be presented over a small volume of space. When many targets are present over a large volume of space, a display capable of hemispherical coverage and of more natural appearance is desirable.

Volumetric Models

Various attempts have been made to simulate the actual position of reflecting objects in a volume by the production of a so-called scale model of space with the target coordinate positions displayed to the same scale.

Two interesting examples of this have appeared recently in the patent literature. The first[3] describes a system, shown in Fig. 7, whereby an ordinary PPI is used to present the output of a volumetric coverage radar. In one embodiment of this invention the ground position of the targets is placed in a storage medium and the height of the target is stored in the medium in a recoverable code. A height decoding mechanism then sequentially scans the storage medium and selects the targets at various height levels (rejecting targets at other height levels). Simultaneously, a mechanism driven from the decoder, positions the tube physically nearer or farther from the observer in synchronism with the height layer being decoded. If the selected layers of altitude and correlated tube positions are selected in synchronism at a fast enough rate, the targets will appear to be situated in a scale model of space at the proper spatial coordinates as derived from the radar.

The second system[4], as shown in Fig. 8, depends for its operation upon a special volumetric chamber to produce a spatial scale model indicator. This chamber has attached to it two ordinary cathode ray guns at right angles to each other. The two electron beams are accelerated into and may intersect within the central chamber. The central chamber is filled with a gas having properties such that it emits light when the beams intersect with proper energy. Light is emitted, however, only at the point of intersection. The proper beam energy for light emission at the crossing point exists only when both beams are energized with target signals simultaneously.

Thus, if the two beams are made to scan the volume of the chamber in such a manner that their intersection represents the point in space being traversed by the radar search system, the chamber will show bright spots at the proper spatial positions. The observer can then interpret the spots of light in the chamber as the three-dimensional position of the reflecting objects in space.

Other means of producing a three-dimensional space display have been investigated. One promising means lies in the principles of stereoscopics.

Stereo Oscillography

The Sperry Gyroscope Company was responsible for much of the early investigation and development of stereo oscillography and was the first company to develop [5,6,7] a method of producing a three-dimensional radar display by means of a volumetric radar and the utilization of stereoscopic principles with cathode ray tubes. In order to grasp the principles involved in such displays it is necessary to understand some of the fundamentals of three-dimensional vision or "depth perception".

How the Eye Sees Three Dimensions

The sensation of depth comes about from the subtle blending of many clues obtained by the eyes in the process of viewing objects. A single eye can obtain the effect from the everyday experiences of looking at objects and noting:

1. Their relative sizes
2. The way the objects appear in perspective
3. The position of one object in front of another
4. The changes in focus of nearby objects
5. The parallax effect seen when an observer shifts his position relative to near and far objects
6. The effect of shadows

A much more pronounced effect is possible, however, when both eyes are used to view objects. This process, known as binocular vision, accomplishes depth perception by the following additional means:

1. Convergence - the angular change in the visual eye paths for varying depth
2. Accommodation or focus - as related to convergence
3. Retinal disparity - the difference in the aspect or backgrounds of the two eye images when the eyes are fixed upon a single object.

From repeated experience the eyes condition themselves to accept the clues given to them by all the processes above and to deduce from the combination the correct spatial position of the objects seen.

It is possible, however, to construct a visual picture which deletes some of the clues. If the eyes do not pick up false clues that generate conflict in interpretation, the missing clues are sometimes unimportant and the eyes continue to see depth.

Stereo Indicator

If there were no clues to depth for the single eye and none for the binocular vision of both eyes, except for the clues of convergence and retinal disparity, it would be possible to present a three-dimensional picture in terms of two flat pictures.

If a typical radar PPI CRT is used to present target data of azimuth and ground range, a perfectly "flat" two-dimensional indicator will result. The light areas on the tube are essentially "structureless" and of the same size, and usually no shadows and no intersections or parallax effects are noted. When such an indicator is viewed, there are no clues available to indicate any third dimension or height of reflecting objects. All objects will appear in the plane of the scope face.

If two identical PPI's of the type mentioned above are arranged so that the right eye sees only the right tube and the left eye sees only the left tube in such a manner that the two images fuse, a binocular convergence effect can be gained by adding a "stereoscopic displacement" to the two scopes. A third dimension is then presented to the operator to enable him to decide the position of reflecting objects in space within the range of the radar.

That such a binocular convergence creates a third-dimensional effect has been shown by Schmitt[8] in the following brief analysis of the stereo displacement. The operator is assumed to be looking down on the objects to be presented, as shown in the diagram of Fig. 9.

S = interocular distance

d = distance from eye to projection plane

c = apparent distance of object from projection plane

Δ = amount of displacement necessary to make object appear at distance $d-c$ from observer

The necessary amount of displacement of each object in order for the object to appear at the distance $d-c$ from the operator or at a distance c from the projection plane is:

$$\Delta = \frac{Sd}{2(d-c)}$$

This relation can be derived from the similar triangles as seen in Fig. 9.

Hence, for an object a given distance in front of the projection plane the electrical displacement (Δ) should be applied with different polarity to the scopes; the right scope should have its displacement to the left and the left scope should have its displacement to the right.

The relation between apparent distance in front of the projection plane and the required displacement is not a straight line function, but over certain limits a straight line function will give sufficiently good results.

Simplified Stereo Indicator Circuit

A representative radar indicator system providing a stereo presentation is shown in Fig. 10. A pencil beam radar scanning antenna, as it is moved in azimuth and elevation, scans a volume of space surrounding the radar location. The elevation and azimuth of the radar scanner are utilized in combination with the time of return of the radar video from the reflecting object, with the combination producing the stereo indication according to the principle outlined above.

A fixed a-c voltage is first applied to the elevation resolver rotor, which moves in direct connection with the position of the scanner elevation. Two outputs are derived from the scanner elevation resolver; they are an elevation sine function and elevation cosine function.

The position of the elevation resolver rotor is chosen with respect to the

scanner elevation so that an a-c voltage whose amplitude is proportional to the cosine of the elevation angle is fed to the ground plane channel and a sine function is fed to the stereo channel.

The sweep former in the ground plane channel is used in combination with the synchronizer trigger to produce a sweep voltage starting at the time of the transmitter pulse and having an amplitude directly proportional to the cosine of the elevation angle. Therefore, the effect is to shorten the sweep length with increasing elevation or, in other words, to present a true ground projection of objects which lie in the three-dimensional space.

The cosine function sweep is applied to the rotor of the azimuth resolver. The rotor of the azimuth resolver is moved in direct connection with the azimuth position of the radar scanner. The output of the resolver stators is two sweep voltages which vary in amplitude 90° out of space phase with each other as the scanner rotates.

These two azimuth resolver outputs are fed into the inputs of horizontal and vertical PPI amplifiers which feed the horizontal and vertical plates of the left and right CRT. The horizontal and vertical amplifiers are connected to produce push-pull voltages from the single ended horizontal and vertical stator inputs. Thus, as the scanner rotates, a ground plane PPI will be generated by sweeps on both right and left CRT scopes, which are identical and which represent the horizontal projection of the radar scanning point in space.

The manner in which the stereo displacement is produced is as follows: the other stator winding of the elevation resolver produces an a-c voltage whose amplitude is proportional to the sine of the elevation angle of the scanner. In a manner identical to that of the ground plane (cosine) channel already discussed, the stereo channel sweep former produces a sweep voltage starting at the time of the transmitter trigger and having an amplitude which is a sine function of the elevation angle. Therefore, there is applied to the phase splitter a sweep voltage which is a function of the altitude of the radar scanning point.

The phase splitter produces sweeps of equal amplitude but of opposite polarities. One sweep is fed to the horizontal amplifier for the left CRT and its inverse is fed to the horizontal amplifier for the right CRT. The polarities are such that the displacement is to the right for the left scope and to the left for the right scope. The radar targets thus appear to be in their true three-dimensional position in space except for the aforementioned distortion caused by displacement non-linearity as related to altitude.

Fig. 11 illustrates the type of pattern obtained from the left and right scopes of the radar stereo indicator described above. Careful examination of these patterns discloses that certain reflecting objects have the <u>same</u> scope face position on the two scopes. These are objects that lie in the ground plane or plane of the scopes. Corresponding images of other objects have been displaced horizontally to the right on the left scope and to the left on the right scope. These are objects which will appear <u>above</u> the plane of the scopes toward the observer when the two scope images are viewed stereoscopically.

In the stereo displacement diagram, Fig. 9, the displacements shown were sufficient to move the object from the remote position beyond the scope face to a position in front of the scope between observer and scope face. In most practical systems (including the one described) the most remote objects are usually shown in the plane of the scope face and hence no electrical deflection is needed to bring the objects from the remote point to the plane of the scopes.

<u>Methods of Combining the Scope Images</u>

Various methods of combining the scope images are available:

1. Direct eye-paths - If the scopes are arranged so that they can be viewed directly by a single observer the paths can be physically separate, as illustrated in Fig. 12a. When large cathode ray tubes are used, the mirrors are necessary to insure that a reasonable convergence distance to the scope face can be realized.

2. Lens system - By the use of a lens system and small cathode ray tubes as shown in Fig. 12b it is possible to avoid the use of mirrors and to space the tubes at about interocular spacing. This is similar to the system used in the viewing of slides used in ordinary stereoscopes.

3. Polarized viewing[9] - A half-silvered mirror (one specially treated to transmit and reflect light equally) can be used to fuse the two images as shown in Fig. 12c. The combination of polarizing filters placed in front of the tube and polarizing glasses worn by the observer accomplishes the separation of the eye paths. By proper orientation of the separate eye pieces in the goggles in relation to the filters in front of the tubes, good light transmission is obtained in relation to one CRT and almost complete light

cancellation is obtained in relation to the other CRT.

Various other methods can be used to accomplish the stereoscopic principles of displacement for the separate eye images, fusion of these images, and separation of the left and right eye paths [10,11].

Precautions To Be Observed in a Stereoscopic System

The stereo method outlined above depends for its operation upon the eyes' properties of convergence and retinal disparity previously defined. Since all scope indications are at one focus, however, the focusing properties of the eyes are not used in the indication of the third (depth) dimension. Precautions must be taken to insure that the focus effect does not confuse the eyes and cause strain.

Since the principal focus changes of the eyes occur within two yd[12], the stereo system should be arranged such that all the apparent visual distances to the indicated objects are greater than two yd from the observer. This prevents the conflict that arises from permitting focus to tell the mind that the object is at one range and convergence (or retinal disparity) to tell the mind that the object is at another range.

Both binocular convergence and retinal disparity depend upon the difference in the pictures presented to the left eye and right eye in order to detect depth. These two clues for detecting depth are effective to about 20 yd in range. Therefore, no object should be displaced in range so as to appear to be greater than 20 yd from the observer. This defines the maximum viewing field for the third (depth) dimension. The optical system and stereo displacement should be such as to cause convergence of the eyes on points between two and twenty yd, and the accommodation arrangement should be such as to cause the eyes to focus at some midpoint such as ten yd.

In general, the amount of convergence should be limited by keeping parameter distances within a range somewhat smaller than 18 yd. This is especially true where many objects must be observed: detecting depth by retinal disparity is both difficult and tiring if the convergence of the eyes varies appreciably from an object to its background.

The field of view of the display should be such that the smallest indication on the scope face should be well within the resolving power of the eyes. The resolving power or separation threshold between two dots visible to the human eyes is approximately 3 min of arc[12]. If one assumes about 1000 lines as the definition of an ideal CRT (about twice as good as the best of present-day scopes) and that such a CRT be used in combination with a radar capable of this range resolution, the field of view should be around 50° (3000 min).

The optimum apparent viewing parameters have thus been specified and are shown in Fig. 13.

Depth Resolution

Depth resolution is defined as the smallest increment in depth which can be detected by the eyes when viewing the display. The amount of resolution in depth can be found by considering the minimum detectable changes in the principal depth clue - that of convergence. The human ocular system is capable of detecting surprisingly small angular changes in convergence. (Lowest threshold is two sec and the average is twelve sec of arc[12].) The amount of depth resolution, therefore, may be found for the stereo display of Fig. 13 by dividing the angular difference in convergence between the nearest (2 yd) and farthest (16 yd) object by the average angular convergence threshold. The result is that it is possible to detect about 450 visible stereo levels. However, as a practical matter, the assumption of a maximum radar height detection equal to radar range (scope radius) reduces this figure to approximately 50 levels.

Results of Tests of Stereo Perception of Human Operations on a Simulated Display

In order to test the general effectiveness of the stereo presentation under ideal conditions, a stereo console was constructed as part of a specific study of indicator techniques. The console utilized the optical arrangement shown in Fig. 12c. The scope patterns shown in Fig. 11 were reproduced with fluorescent paint and were illuminated with ultraviolet light from shaded lamps. One hundred and ten subjects were tested (engineers, technicians, and secretaries) with no preparation as to what they were expected to see. They were placed before the console, their polaroid glasses were adjusted, and they were asked to explain what they saw. One hundred and three subjects were able to see the targets in three dimensions. Of the seven who did not see the effect only one had 20-20 vision. The rest had varying degrees of visual defects.

In general, the subjects had no trouble in picking the correct relative heights of the targets presented and were able to estimate the exact height of the targets within about 20 per cent. (Since the exact apparent height will vary with

viewing distance of the eyes from the display, a height scale should be added where exact heights are desired.)

Experimental Stereo Indicator

A radar PPI indicator utilizing the stereo principles has been built and tested by Sperry as part of the development of a search radar. The system utilized an optical arrangement as shown in Fig. 12a with the addition of a magnification system to increase the apparent size of the 3 in. diameter scopes used. The indicator displayed in three dimensions the volume of coverage of the radar system. The experiments afforded much basic information useful in future work and they brought out certain difficulties as described below.

Disadvantages of Stereo System

The test program previously described has demonstrated that the stereo system has certain disadvantages.

Operational Disadvantages:

1. The problem of eye strain in a stereoscopic system is often more serious than it is in a "flat" display.

a. Incorrect optical parameters may give false clues which confuse the eyes and lead to eye strain.

b. Noise and irregular clutter voltages which appear on the display may give false clues and may tend to confuse the eyes.

c. In a direct view system, such as is shown in Fig. 12a and b, the requirement that the head be fixed in order that the eye paths remain separate may lead to fatigue.

d. In a polaroid system such as is shown in Fig. 12c, the addition of the glasses and the consequent loss of available light may be tiring.

2. In a radar stereo display system presenting only a few depth clues, the problem of interpretational ambiguities can be troublesome. For instance, two reflecting objects which are situated in space in the same horizontal plane (same h position in an x, y, h coordinate system) and which are close together (same y position, adjacent x positions) can be interpreted on the stereo indicator as being two reflecting objects at the same ground position (same x, y position) but at different heights. Such confusion of position is caused by the inability of the observer to determine which of the two adjacent objects seen by one eye should be paired with which of the adjacent objects seen by the other eye. Other clues, however, such as persistence of target path, may be sufficient to remove the ambiguities.

3. The persistence of present day scope phosphors is too short to take advantage of volumetric radar coverage. By the time a pencil beam radar scanner has covered a significant part of the total volume, the scope indication representing many of the reflecting objects has disappeared and, therefore, the volumetric indication is not complete.

Mechanization Difficulties:

1. The requirement for two cathode ray scope patterns which are identical (except for stereo displacement) places a severe requirement on cathode ray tube characteristics as well as on the circuits. A departure of more than three or four per cent from similarity adds confusion and decreases the effectiveness of the three-dimensional properties.

2. The requirement for visual fusion of the two images coupled with the necessity for correct viewing parameters requires a rather elaborate optical system and therefore adds significantly to the cost of the system.

Application of New Techniques

Some of the disadvantages to the stereo type of indicator system eventually may be overcome by the use of special display tubes which may replace the conventional cathode ray tubes.

Improved direct view storage tubes[13,14] should furnish superior indicator display performance as follows:

1. The light output from the tube screen can be many times that of a conventional cathode ray tube.

2. The storage characteristic can be used to give an image that does not decay between scans, i.e., ideal persistence characteristics are possible.

3. The storage characteristic may be used to give considerable signal to noise improvement[15].

A direct view storage tube stereo system which is capable of sufficient light output and persistence to present a steady, bright display with superior signal to noise characteristics may eventually be developed. Such a display would be free of many of the disadvantages mentioned in the previous section.

Advantages and Applications

The stereo type of display has two marked advantages over other more conventional displays which present three-dimensional data:

1. The display has an appearance of naturalness. It presents a "scale model of space" rather than a synthetic representation.

2. The display can be interpreted more rapidly than many conventional displays. This is especially true in cases where there are large numbers of reflecting objects in the volume of space covered by the radar.

These two advantages are most important in many surveillance and search type radar systems in which three-dimensional information concerning a large number of reflecting objects is necessary.

Typical Application

Airport surveillance radar is one type of equipment which might benefit from the type of display provided by a stereo indicator. At present altitude information is not automatically furnished by the usual airport surveillance radar and altitude data must be obtained from the aircraft. This is generally slow and places an unnecessarily heavy load on the communication system[16]. A radar supplying three-dimensional data to a stereo indicator might fulfill the requirements of this application.

Conclusion

The stereo method of indicating three-dimensional data, while having some definite disadvantages, holds considerable promise when proper design can be combined with superior indicating tubes now under development.

Acknowledgement

The author wishes to acknowledge the assistance and suggestions furnished by Mr. Curtis Jansky of the Ground Armament Engineering Department of the Sperry Gyroscope Company. Mr. Jansky has furthered the development of certain stereo indicator techniques mentioned in this paper and been responsible for much of the work on optimum stereo viewing parameters included herein.

Bibliography

[1]Staff of Radar School, Massachusetts Institute of Technology, Principles of Radar, 3rd edition, (McGraw-Hill Book Co., New York and London, 1952).

[2]C. Berkley, "Three-Dimensional Representation on Cathode-Ray Tubes," Proceedings of the I.R.E., (December 1948), pp. 1530 ff.

[3]Patent #2,602,921 Aircraft Traffic Control System, Isbister and Peters (Sperry), July 8, 1952.

[4]Patent #2,604,607, Three-Dimensional Indicator Tube, Howell.

[5]Patent #2,514,828, Synthesized Stereoscopic Vision, W. A. Ayres (Sperry), July 11, 1950. Application dated September 12, 1942 (#458,109).

[6]Patent #2,434,897, Stereoscopic Radio Location Device, W. A. Ayres (Sperry), January 27, 1948.

[7]Patent #2,426,979, Stereoscopic Range Indication, W. A. Ayres (Sperry), September 9, 1947. Application (#478,586) dated March 9, 1943.

[8]O. H. Schmitt, "Cathode-Ray Presentation of Three-Dimensional Data," Journal of Applied Physics, Vol. 18, (September 1947), pp. 819 ff.

[9]Patent #2,501,748, Synthesized Stereoscopic Range Indicator, E. C. Streeter, Jr., (Sperry), March 28, 1950.

[10]H. A. Iams, R. L. Burtner, and C. H. Chandler, "Stereoscopic Viewing of Cathode-Ray Tube Presentations," RCA Review, (March 1948), pp. 149 ff.

[11]E. Parker and P. R. Wallis, "Three-Dimensional Cathode-Ray Tube Displays," Journal of the Institution of Electrical Engineers, Vol. 95, Part III, (September 1948), pp. 371 ff.

[12]Tufts College Institute for Applied Experimental Psychology, Handbook of Human Engineering Data, Technical Report SDC 199-1-2, 1951.

[13]M. Knoll and B. Kazan, Storage Tubes and Their Basic Principles, Part V, (John Wiley and Sons, Inc., New York, 1952).

[14]S. T. Smith and H. E. Brown, "Direct-Viewing Memory Tube," Proceedings of the I.R.E., (September 1953), pp. 1167 ff.

[15]J. V. Harrington and T. F. Rogers, "Signal-to-Noise Improvement Through Integration in a Storage Tube," Proceedings of the I.R.E., (October 1950), pp. 1197 ff.

[16]R. Hotz, "Radar Breaks Bad-Weather Jams," Aviation Week, (March 16, 1953), pp. 14 f.

Additional References

D. M. MacKay, "Projective Three-Dimensional Displays," Electronic Engineering, (July 1949), pp. 249 ff, (August 1949), pp. 281 ff.

J. V. Harrington, "Storage of Small Signals on a Dielectric Surface," Journal of Applied Physics, Vol. 21, (October 1950), pp. 1048 ff.

W. D. Wright, "The Role of Convergence in Stereoscopic Vision," Proceedings of the Physical Society, (April 1951), Section B, Vol. 64, Part 4, pp. 289 ff.

A. W. Judge, Stereoscopic Photography, (American Photographic Publishing Company, Boston, 1926).

Patent #2,433,002, Pulse Type Position Indicator, December 23, 1947.

Patent #2,440,777, Tridimensional Radio Direction Finder, May 4, 1948.

Patent #2,432,101, Indicating Method and Apparatus for Detection Systems, December 9, 1947.

Patent #2,467,319, Unitary Range, Azimuth, and Elevation Alignment Indicator for Radar Systems, K. L. King, April 12, 1949.

Patent #2,486,197, Three Dimensional Proximity Indicator System, Newbold, October 25, 1949.

Patent #2,513,962, Directive Radiant Energy Object Locating System, Patterson, July 4, 1950.

Patent #2,602,923, Stereoscopic System for Three Dimensional Location of Aircraft, Frazier, July 8, 1952.

Patent #2,544,624, Three Dimensional Oscillograph, Whittaker, March 6, 1951.

Patent #2,538,800, Stereoscopic Radar, Ranger, January 23, 1951.

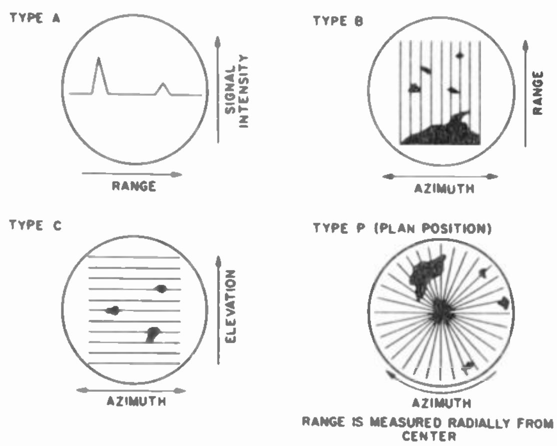


Fig. 1 - Basic indicator types.

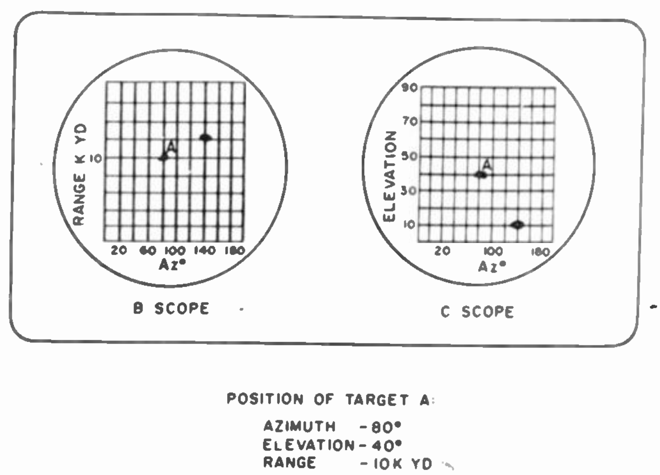


Fig. 3 - Basic indicator types.

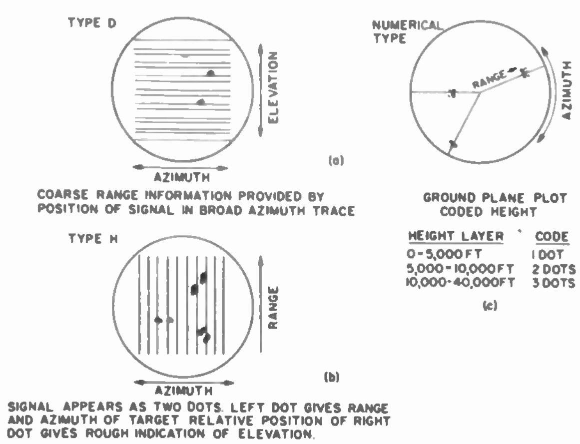


Fig. 2 - Combined indicator.

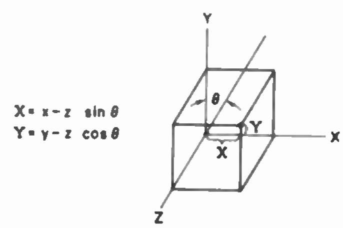


Fig. 4 - Oblique view of cube.

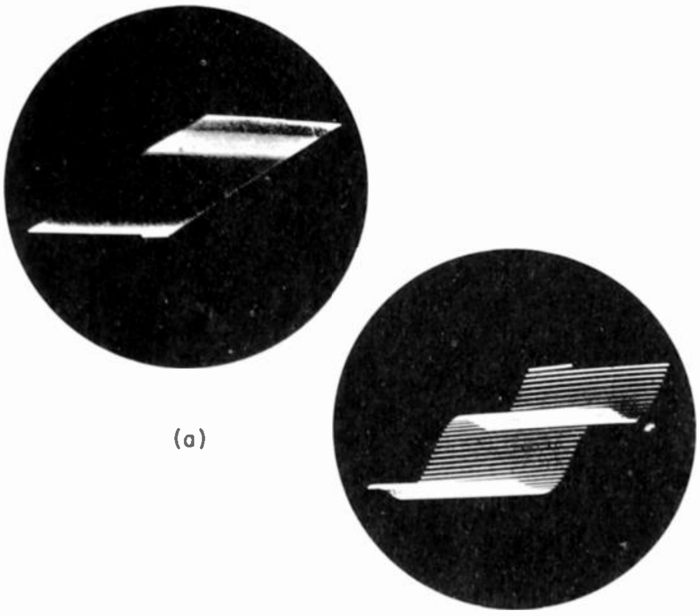


Fig. 5 - Perspective 3-D display.

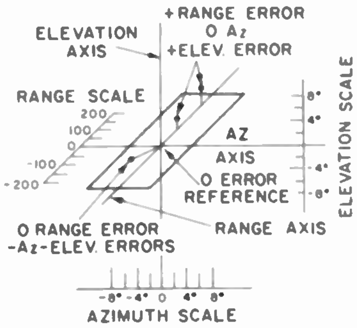


Fig. 6 - 3-D tracking scope.

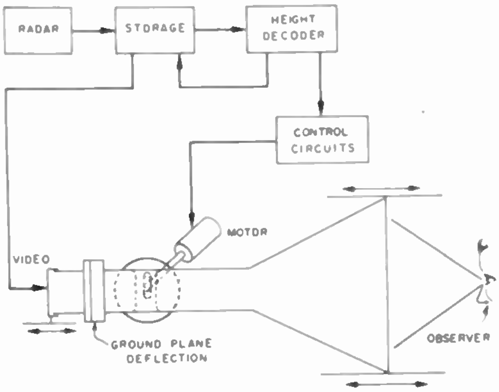


Fig. 7 - Volumetric indicator.

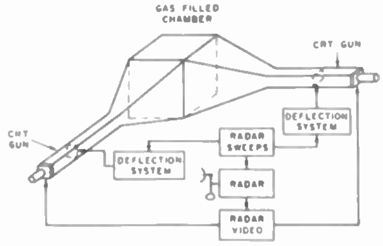


Fig. 8 - Volumetric cathode-ray tube.

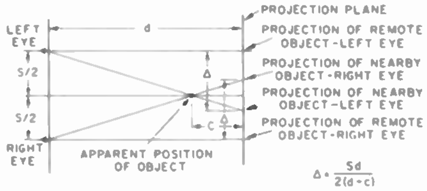


Fig. 9 - Stereo displacement.

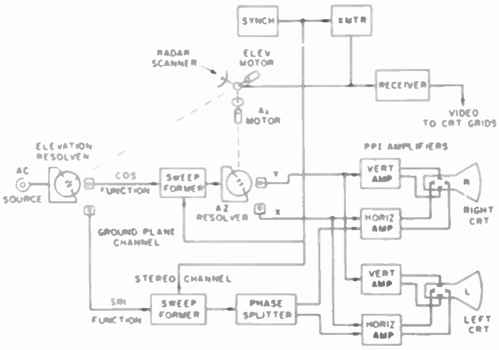


Fig. 10
Simplified block diagram -
radar PPI stereo indicator.

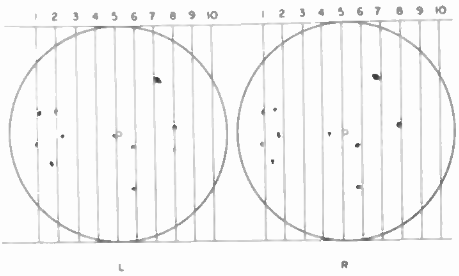


Fig. 11
CRT patterns of radar stereo indicator.

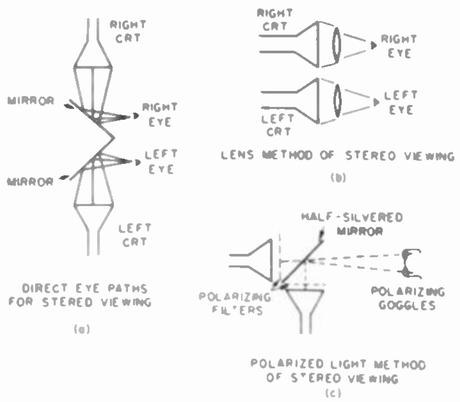


Fig. 12
Polarized light method of stereo viewing.

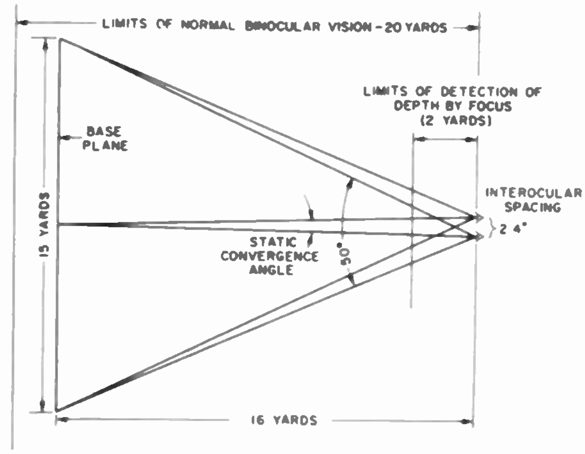


Fig. 13
Optimum apparent viewing parameters of stereo display.

AN AUTOMATIC ANTENNA MATCHING UNIT

E. W. Schwittek
Engineer
Collins Radio Company
Cedar Rapids, Iowa

The theory of operation concerning an automatic antenna matching unit designed to operate in the 2 - 25 mc frequency range is discussed. This device is intended primarily for operation with aircraft transmitters which are designed to deliver power into a 52-ohm resistive load. The purpose of such a unit is to automatically transform the impedances presented by all ordinary aircraft antennas to a 52-ohm resistive impedance for coupling to the aircraft transmitter.

INTRODUCTION

In the design of any practical automatic antenna matching unit, several factors must be considered. The most important of these is that the automatic device must transform the very wide range of antenna impedances commonly encountered in communication antennas to the particular resistive impedance into which the transmitter is designed to operate. The automatic device must perform this function accurately, efficiently, rapidly, and with little or no pre-information regarding the type of antenna and operating frequency.

The particular automatic antenna matching unit which will be described is intended for operation with aircraft antennas in the frequency range of 2 to 25 mc with a transmitter that is designed to deliver its power to a 52-ohm resistive impedance. It is capable of thus transforming the impedances represented by all ordinary aircraft antennas over the 2 to 25 mc frequency range without additional pre-information.

To completely discuss all aspects of the automatic antenna matching problem is beyond the scope of this paper. The servo systems that are employed to drive the tunable elements and the design of high Q tuning elements are subjects for papers in themselves. It is, therefore, the purpose of this paper to (1) relate principally the circuitry and theory of the impedance matching network itself; (2) discuss the operation of the device which supplies the servo system with the information necessary for it to drive the tunable elements to their proper position; and (3) provide a brief explanation of the manner in which all the elements are interconnected to perform the automatic matching function.

THE IMPEDANCE MATCHING NETWORK

The capabilities of the complete automatic matching unit to transform a wide range of antenna impedances is directly dependent upon the capabilities of the impedance matching network itself.

In the particular impedance matching network which is to be described, three very basic matching principals are combined to perform the complete matching function. The first of these, and the most obvious, is that an antenna impedance which contains both resistance and reactance may be made to appear purely resistive by series resonating the antenna. That is, an antenna which has the impedance r + jx may be made to appear as r + j0 by inserting in series with it a reactance equal to the antenna reactance, but of opposite sign.

The second basic matching principal is that exhibited by the "T" section shown in Fig. 1. To more clearly establish a comparison between the circuit elements shown in Fig. 1 and their ultimate representation in the complete impedance matching network, let us define them. The element r, denotes the antenna resistance after the antenna reactance has been cancelled according to the first matching principal. The element R is representative of the resistive impedance presented to the transmitter by the impedance matching network, X_2 is the reactance of a fixed capacitor, and X_1 and X_3 are variable tuning elements.

If the impedance equation is written for the above circuit, Eq. 1, and set equal to R + j0, it is possible to obtain two simultaneous impedance equations (Eqs. 2 and 3) in which r and X_3 are variables.

$$R + j0 = \frac{r X_2^2}{r^2 + (X_2 + X_3)^2} + j\left[\frac{X_2(r^2 + X_2 X_3 + X_3^2)}{r^2 + (X_2 + X_3)^2} + X_1\right] \quad (1)$$

$$r^2 + X_3^2 + r\left(\frac{-X_2^2}{R}\right) + X_3(2X_2) + X_2^2 = 0 \quad (2)$$

$$r^2 + X_3^2 + X_3\frac{(2X_1X_2 + X_2^2)}{(X_1 + X_2)} + \frac{X_1 X_2^2}{(X_1 + X_2)} = 0 \quad (3)$$

Fortunately, these simultaneous equations have the basic form of circles and, therefore, their solutions may be readily obtained by plotting the circles and noting their points of intersection. Such a plot is shown accompanying the circuit in Fig. 1.

It is important to realize exactly how the plot is determined. All variables and constants are expressed in terms of the value of R. That is, they are expressed in terms of per unit values with R as the base. Since it represents a fixed capacity, X_2 can be considered constant at a given frequency. In this particular example, it is considered to have the value of $-j4R$. Plotting this information yields the X_2 circle which is the plot of equation (2).

Other circles are drawn on this graph using the second simultaneous impedance equation (Eq. 3) for various values of X_1, wherein the value of X_1 is expressed as a per unit value of R.

The locus of the intersections of the circles, indicated by the solid line, represents the solutions of the simultaneous impedance equations for various values of X_1. This plot, therefore, reveals that any value of r, antenna resistance, on the locus of solutions may be transformed to R, the required resistive load for the transmitter, with appropriate values of X_1 and X_3. If R is 52 ohms, as we will require for the example transmitter, then this circuit is ideally suited for transforming antenna resistance values greater than 52 ohms to 52 ohms.

The third basic matching principal is that exhibited by a "T" section also, but under slightly different conditions. The circuit employed for the third matching principal is shown in Fig. 2.

Again, r represents the various antenna resistances desired to be transformed to R, the required transmitter load. Reactance, X_3, is a variable tuning element, and X_1' and X_2' are tuning elements arranged so that the sum of X_1' and X_2' is a constant for any particular frequency.

Since the circuit is basically a "T" section, the same type of equations are used for the solution of the third matching principal as was used for solution of the second.

$$R+j0=\frac{r(X_2')^2}{r^2+(X_2'+X_3)^2}+j\left[\frac{X_2'(r^2+X_2'X_3+X_3^2)}{r^2+(X_2'+X_3)^2}+X_1'\right] \qquad (4)$$

$$r^2+X_3^2+r\left[\frac{-(X_2')^2}{R}\right]+X_3(2X_2')+(X_2')^2=0 \qquad (5)$$

$$r^2+X_3^2+X_3\frac{[2X_1X_2'+(X_2')^2]}{(X_1+X_2')}+\frac{X_1(X_2')^2}{(X_1+X_2')}=0 \qquad (6)$$

Again, the simultaneous equations (5 and 6) exhibit the general form of circles. The differences between the second and third basic matching principal is due primarily to the fact that X_2' is now an inductance rather than a capacity, and that though now both X_1' and X_2' are variable tuning elements, their sum is a constant.

As before, it is important to realize how the plot is established. In this example $X_1' + X_2'$ is a constant and is set equal to $+j4R$. Graphical simultaneous solutions are obtained by first assuming that $X_1' = +j1R$ and $X_2' = +j3R$, then $X_1' = +j2R$ and $X_2' = +j2R$, etc. The intersections of the various circles again produce a locus which represents the solutions of the simultaneous impedance equations for various values of X_1' and X_2' within the limit that the sum of X_1' and X_2' is a constant. This plot, therefore, reveals that any value of r, antenna resistance, on the locus of solutions may be transformed to R, the required resistive load for the transmitter, with appropriate values of X_1', X_2' and X_3. Again, if we desire R to be 52 ohms, this circuit is ideally suited for transforming antenna resistance values which are less than 52 ohms to 52 ohms.

The circuits of Fig. 1 and 2 may be very simply combined with the aid of a device, hereafter referred to as an autotransformer, to produce a new circuit that is capable of transforming resistance both above and below 52 ohms to 52 ohms. A schematic of this combination circuit is shown in Fig. 2.

With reference to Fig. 3 it may be seen that the autotransformer is a coil with a fixed tap and a variable tap. When antenna r values greater than 52 ohms are desired to be transformed to 52 ohms, the variable tap is in some position above the fixed tap of the autotransformer. The matching qualities of the second basic matching circuit are thus obtained. It is obvious that some variation in the theory must exist due to the presence of the portion of the autotransformer from the fixed tap to ground, however, if $X_1' + X_2'$ is large compared to R it does not materially affect the basic matching theory.

On the other hand, when antenna r values less than 52 ohms are desired to be transformed to 52 ohms, the variable autotransformer tap is in some position below the fixed tap of the autotransformer, as indicated by the dotted position of the variable tap. The X_1 portion of the autotransformer is eliminated from the circuit and X_1' and X_2' appear. Again it is obvious that the presence of X_2 in this situation tends to produce some variation in the matching theory presented by altering the value of $r - jX_3$ which it shunts. However, a slight readjustment of X_3 and the variable tap position of the autotransformer from the theoretical will cancel its effect. The presence of X_2 complicates the circuit slightly, but does not alter its basic operation.

The mutual inductance existing between X_1'

and X_2' has been neglected in the foregoing circuit analysis. Its only real effect would be that of altering the assumption that $X_1' + X_2'$ is a constant at all variable tap positions which are below the fixed tap. This is not particularly serious, however, since at the limits when $X_1' = 0$ and $X_2' = +j4R$, or when the reverse is true, the mutual inductance is zero. Therefore, the end points of the locus of matchable values of r (Fig. 2) would not be altered. The only effect would be that the shape of the locus between its end points would deviate slightly from what is shown.

The complete radio frequency matching network is shown in Fig. 4. It differs from the previous circuit in that a variable inductance as well as a variable capacitance is inserted in series with the antenna. These are the reactive elements used to series resonate the antenna. The variable capacitor serves the dual function of series resonating any inductive antenna and providing X_3.

It is felt that the circuit of Fig. 4 does satisfactorily perform the required function of transforming a very wide range of antenna impedances to 52 ohms resistive.

THE DISCRIMINATOR

The source of information that is used to direct the operation and ultimate tuning position of the various r-f matching network components is an impedance monitoring device which will hereafter be referred to as a discriminator. As Fig. 5 indicates, the discriminator is connected between the transmitter and the r-f matching network and performs the function of constantly monitoring the impedance presented to the transmitter. The impedance monitoring is done with respect to the ideal $52\underline{/0^\circ}$ ohm transmitter load. That is, the discriminator monitors both the magnitude and phase of the impedance presented to the transmitter and submits its measurement in terms of two d-c voltages which have polarity dependent upon the direction of impedance deviation from $52\underline{/0^\circ}$ ohms. Let us first examine the portion of the discriminator that is sensitive to the phase angle of the impedance presented to the transmitter.

The phase of the impedance presented to the transmitter by the r-f matching network is monitored by noting the phase relationship between the transmission line voltage, E, and the line current, I. To determine the phase relationship existing between E and I it is, of course, necessary to sample each of them. The means of sampling is important since it must be accurate throughout the wide frequency range over which the system is required to operate.

The line current, I, is sampled by inductive coupling to the transmission line. The center conductor of the coaxial transmission line is passed through the center of a 5/8 inch diameter powdered iron ring. A center-tapped, five-turn torroid winding about the ring serves to provide the inductive coupling to the line, wherein the induced voltage produced in the five-turn winding is proportional to line current and 90 degrees out of phase with it. The line voltage, E, is sampled by a capacitive divider consisting of C_1 and C_2.

It may be seen that the means of sampling E and I has provided the essential voltages at the proper phase so that they may be added vectorially in the same manner as that employed in the ordinary Foster-Seely discriminator circuit. That is, if $\vec{E_1 + E}$ is rectified and added in polarity opposition to rectified $\vec{E_2 + E'}$, zero d-c output voltage is obtained if E and I are in phase. If E and I are not in phase, the circuit will provide d-c output voltage of a polarity dependent upon whether E leads or lags I. With this type of E and I sampling it is practical to obtain phase monitoring of E and I to within ± 5 degrees over the 2 - 25 mc frequency range.

Let us now examine the portion of the discriminator that is sensitive to the magnitude of the impedance presented to the transmitter. It was previously stated that the induced voltage across the five-turn torroid winding is directly proportional to the magnitude of I. It is also directly proportional to frequency since the voltage drop across the portion of the line to which the torroid is coupled (primary) is directly proportional to frequency.

The same proportion properties can be made to exist between E_3 and E through the use of a simple circuit containing C_3 and R', when $R' \ll X_{C3}$.

With the frequency denoted by f, the magnitude of $E_4 \propto K_1 fI$. The magnitude of $E_3 \propto K_2 fE$. Through the choice of component values it is possible to make K_1 and K_2 constants which will allow $E_3 = E_4$ when $E/I = 52$ for all frequencies. Then, if rectified E_3 is added in polarity opposition to rectified E_4, zero d-c output voltage is obtained from this portion of the discriminator when $E/I = 52$. If this ratio is not true, the discriminator will provide d-c output voltage of a polarity dependent upon whether the ratio of E to I is greater than or less than 52. With this type of E and I sampling it is practical to obtain ratio monitoring of E and I to within $\pm 10\%$ of 52 over the 2 - 25 mc frequency range.

The complete discriminator is then a device with two separate d-c voltage outputs. When the impedance presented to the transmitter is $52\underline{/0^\circ}$ the d-c voltage from each output is zero. When the impedance presented to the transmitter differs from this value, the d-c outputs are not zero and have polarity dependent upon the direction of magnitude or phase deviation.

OPERATION OF THE IMPEDANCE MATCHING NETWORK FROM DISCRIMINATOR INFORMATION

To more clearly explain the manner in which the impedance matching network operates from discriminator information, let us analyze their combination with the servo system as shown in

Fig. 6. The elements C and L are termed phasing elements. Together they provide the large range of reactance which is necessary to series resonate the antenna reactance and provide the X_3 reactance that is essential to the impedance matching properties of the "T" section. The reactance of this element would be such that $x + X_L$ or $C = X_3$. Obviously, for any particular tuning operation, only one of these reactive elements is necessary. This is accomplished in the actual unit by control circuits which retain C at its maximum value when L is required for tuning and L near its minimum value when C is required for tuning.

Examination of Fig. 6 indicates that the chopper coils and the fixed phases of the servo motors are all excited from a common 400 cycle source. The servo amplifiers themselves are merely 400 cycle amplifiers, which in addition to amplification provide a 90-degree phase shift to the signal that is passed through them.

As an example of operation, let us assume that the various impedance matching network tuning elements are in such a position that only a slight discrepancy exists between the actual impedance presented to the transmitter and the required impedance of $52\underline{/0^{\circ}}$ ohms. For the sake of explanation, assume that this impedance is $60\underline{/10^{\circ}}$.

At the instant the transmitter delivers r-f energy, both the impedance sensitive and phase sensitive portions of the discriminator will detect this error from the ideal impedance and will each provide a d-c voltage of a particular polarity at their output point. The d-c voltage at the output of the phase sensitive discriminator is applied to a 400 cycle chopper which converts the d-c voltage to 400 cycle a-c having phase in accordance with the polarity of the d-c voltage. The 400 cycle a-c is then amplified, shifted 90 degrees in phase, and is used to excite one phase of a two-phase motor. With the particular 10 degree impedance phase angle error that was assumed, the motor would operate the phasing element, L or C, to increase the value of capacitive reactance in the phasing circuit and reduce the phase angle of the impedance presented to the transmitter to zero degrees.

At the same time, the d-c voltage appearing at the output of the impedance sensitive discriminator is applied to another chopper, servo amplifier and two-phase motor which mechanically moves the variable autotransformer tap nearer ground to correct for the assumed 8-ohm error in impedance magnitude.

This is the basic operation of the system, with one exception. As was stated, the preceding explanation of operation was indicative of a tuning sequence wherein the matching elements were only slightly mistuned from their ultimate positions. Upon further examination of the circuit, it becomes apparent that a more complex condition can, and usually does exist.

For an explanation of this, let us assume that the antenna phasing elements, L and C, are considerably mispositioned from their ultimate tuning points. Under these conditions very little antenna current is able to flow. In fact, most of the r-f current which is being sampled by the discriminator is the inductive current flowing from the fixed tap of the autotransformer to ground or the current flowing from the fixed tap through the capacitor, C_2, to ground, depending upon the starting position of the autotransformer variable tap. The obvious problem exists - that of trying to correctly tune the phasing elements, L and C, to series resonate the antenna with information obtained from a discriminator, which under these conditions cannot detect antenna current because of its relative absence.

This is actually the usual tuning condition and results in the necessity of forcibly operating the phasing elements at the beginning of each tuning operation by automatically applying a voltage to their motor until such time as they are near enough to resonating the antenna to allow the discriminator to assume control of the situation and complete the tuning function.

Due primarily to practical considerations with regard to component minimum and maximum values it is still possible to present an antenna to this network, as it is thus far described, which the network is not capable of transforming to $52 + j0$. Since this cannot be tolerated, in these instances a relay operated fixed capacitor, Cs, provided within the antenna matching unit, is permitted to shunt the antenna and change its effective impedance to a matchable value. Whenever Cs is required for the solution of a matching problem, it decreases matching efficiency. Its use, therefore, is restricted to antenna impedances which represent limits of tuning range and for which the construction of a series resonating reactance or autotransformer would be impractical. The necessity for use of Cs is determined automatically by providing switches at the extreme limits of L and C and the autotransformer, such that if any one of these elements reaches the limit of its tuning range, it will operate the relay which connects Cs in shunt with the antenna.

CONCLUSION

Let us briefly review the original requirements of wide range impedance matching, tuning accuracy, efficiency, rapid tuning and pre-information as they apply to this particular automatic antenna matching unit.

The unit as constructed is capable of operation with antennas from 25 to 90 feet in length over the 2 - 25 mc frequency range. The antennas may have their remote ends either grounded or ungrounded.

The tuning accuracy of the entire unit is approximately equivalent to a 1.3/1 standing wave ratio.

The efficiency of this antenna matching unit is dependent primarily upon the relationship

existing between the antenna Q factor and the Q factors of the phasing inductance, L, and autotransformer. The greater the value of coil Q with respect to antenna Q, the greater is the matching efficiency of this network. Every effort has been extended, therefore, to provide high Q windings for L and the autotransformer.

The time involved in matching any particular antenna is, of course, dependent upon the antenna and the starting position of the various tunable elements. Time values of from 1 to 25 seconds represent the limits in time that will elapse from the instant the transmitter is keyed to the moment the antenna has been completely tuned.

If the antenna length is between 25 and 90 feet and the frequency of operation of the transmitter is between 2 - 25 mc, no further information is required by the unit except a grounding pulse from the transmitter whenever a new frequency is selected. Control circuits within the automatic antenna matching unit perform the function of keying the transmitter prior to tuning, and unkeying the transmitter when the antenna has been properly tuned.

Some changes in antenna characteristics occur during operation, particularly in aircraft, wherein antenna impedance variation occurs between ground operation and flight. The unit is designed to continually monitor the impedance presented to the transmitter and thereby is able to continually correct for this type of variation.

ACKNOWLEDGMENT

The techniques exemplified in this paper represent the efforts of many. In particular, the author wishes to gratefully acknowledge the advice and guidance of D. W. Weber and the contributions of J. Sherwood, S. Morrison, M. Ludvigson and V. Newhouse.

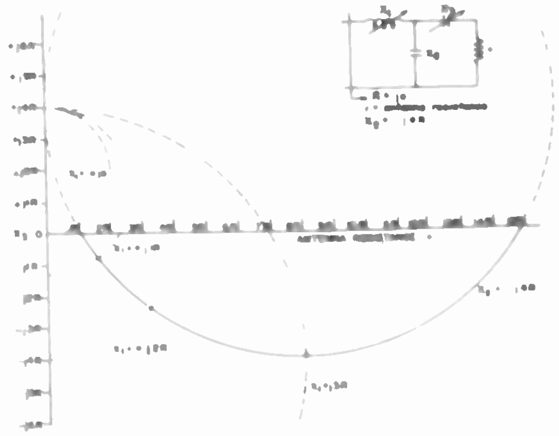


Fig. 1
"T" section matching properties when X_1 and X_3 are variables.

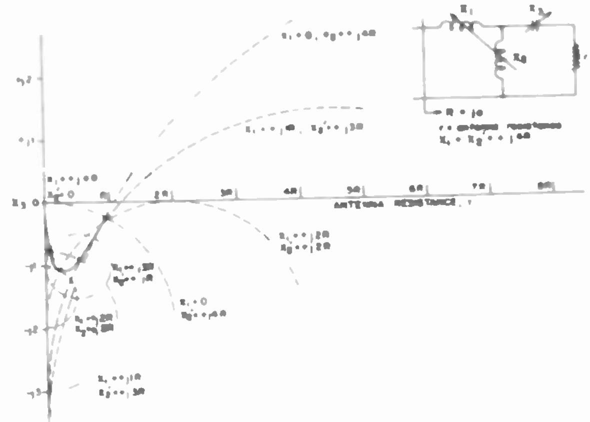


Fig. 2
"T" section matching properties when X_1', X_2', and X_3 are variables.

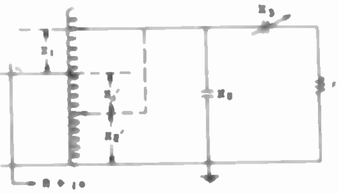


Fig. 3
Use of autotransformer to combine properties of both types of "T" matching sections.

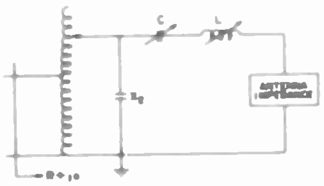


Fig. 4 - Complete impedance matching network.

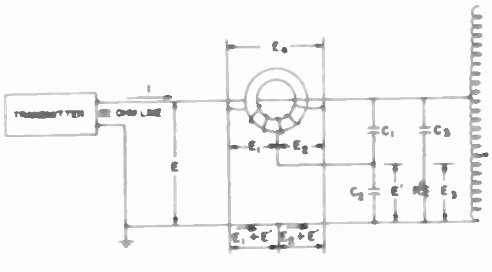


Fig. 5
Phase and impedance sensitive discriminator.

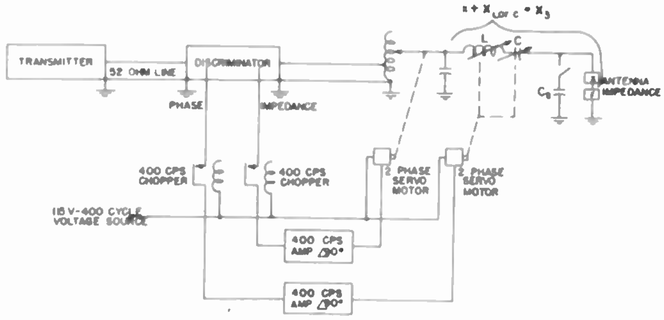


Fig. 6
Block diagram of complete automatic antenna matching unit.

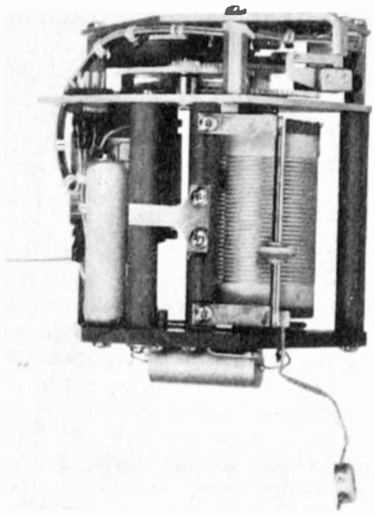


Fig. 7 - Autotransformer assembly.

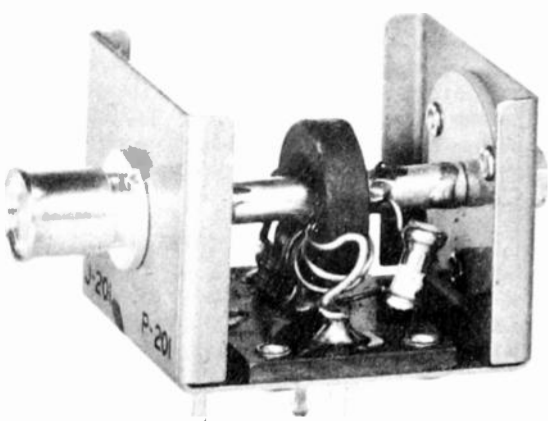


Fig. 8 - Discriminator torroid assembly.

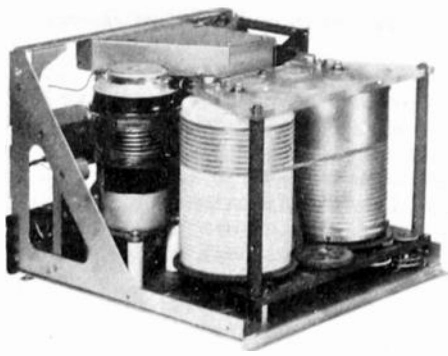


Fig. 9
Rear view of automatic antenna matching unit showing L and C.

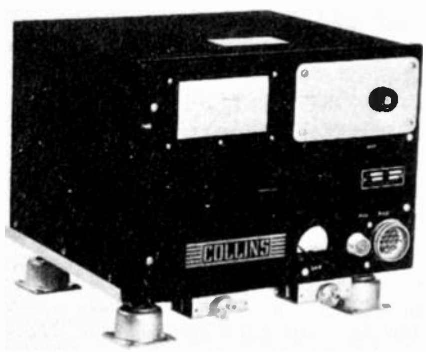


Fig. 10
Complete automatic antenna matching unit.

A PROPORTIONAL DATA TRANSMISSION SYSTEM

W. C. Petrie
Collins Radio Company
Cedar Rapids, Iowa

Summary

A system for radio transmission of continuously variable data is described. The DC input signal is amplitude modulated on a subcarrier which is then transmitted. In order to provide polarity sense and a reference zero for the demodulation process, an unmodulated subcarrier is also transmitted. The system features a stable zero and good linearity with a tolerance for reasonable incidental phase shifts. The dynamic response extends from DC to above 25 cycles per second and the DC to DC gain is limited only by the gain and dynamic range of the AC amplifiers.

Introduction

In a system wherein it is required to transmit continuously over a radio link a number of channels of DC and low frequency proportional data, some form of subcarrier amplitude, phase or frequency modulation is often utilized. Time division multiplex involving digital coding may not be feasible because of the complexity of the required terminal equipment. Use of frequency division multiplex with direct modulation of subcarriers eliminates the necessity of complicated encoding and decoding equipment.

As in any system where DC must be transmitted, the zero drift or stability is of great importance, especially if the system is of the open-loop type. The open-loop case is the usual for the system where data is to be transmitted to a remote point by radio, since the system output is not conveniently available for comparison with the input feedback fashion.

Phase and Frequency Modulation Systems

Where phase or frequency modulation of a subcarrier is used the phase or frequency stability of the system is the major factor affecting the DC stability. Any phase or frequency drift appears directly as an error in the system output. Phase drifts are encountered whenever reference and modulated subcarriers pass through parallel filter networks the drift characteristics of which are not perfectly matched for temperature and frequency variations. In the FM system a spurious component appears in the output whenever the discriminator center frequency and the transmitted subcarrier frequency drift relative to each other.

The obvious cure for this is to design the system in such a manner that all usable input signals frequency or phase modulate the subcarrier to such a degree that drifts are negligible in comparison or are "swamped out" by brute force. This can usually not be accomplished without difficulty, however. If, for example, a system zero stability of better than 1% of the full scale signal is required, and subcarrier phase modulation is proposed, then the phase deviation representing the full scale signal must be at least 100 times the phase stability of the system. The linearity of the phase deviation would of course be specified. A system phase stability of better than several degrees may be difficult to achieve when reference and modulated subcarriers must pass through frequency-selective filters. At the same time the corresponding full scale deviation of several hundred degrees is difficult to generate and demodulate without ambiguity.

For a similar DC stability requirement in a frequency modulated system a similar problem is met. In order effectively to swamp out frequency and discriminator drift, the frequency deviation required to represent the full scale signal may be a relatively large percentage of the center frequency. Modulator and discriminator design are difficult and the available subcarrier spectrum can not be used economically.

Amplitude Modulation System

Brief Description

In the DC transmission system to be described, direct double-sideband suppressed (sub)carrier amplitude modulation is utilized. In order to provide polarity sense and a reference zero for the demodulation process an unmodulated or reference subcarrier is also transmitted. Demodulation is effected by means of a well-known phase detector circuit. It will be shown that the system zero is well protected without undue design difficulties in obtaining reasonable linearity.

Fundamentally, this is the simple approach frequently applied where a DC amplifier is required without the complications of cascaded, direct-coupled stages.

Block Diagram

A simplified block diagram of the system is shown in Figure 1. A subcarrier is generated at the desired frequency in a simple audio oscillator. A reasonably good waveform is desirable although harmonics can be removed by filtering. This signal passes through a frequency doubler after which it is applied to the carrier terminals of a balanced modulator. The DC control signal to be transmitted is applied to the modulation terminals. The output of the balanced modulator is filtered to remove

modulation frequency components and harmonics of the modulated subcarrier.

At this point we have a double-sideband suppressed-carrier signal. Contained therein is information relative to both amplitude and polarity of the input control signal. The subcarrier amplitude is instantaneously proportional to the amplitude of the control signal; and the subcarrier phase is either zero or 180° according to whether the control signal polarity is positive or negative. For a control signal equal to zero the balanced modulator has nominally zero output.

The output of the subcarrier oscillator, taken ahead of the frequency doubler, is also filtered to suppress its harmonic content. This unmodulated signal, at half the frequency of the modulated subcarrier, is then combined linearly with the modulated subcarrier and applied to the modulation input terminals of the radio transmission link. Other subcarrier pairs might be injected at this point to provide additional channels for data transmission.

This composite signal is transmitted as modulation on the carrier of the radio link and is recovered at the radio link output. It is then applied to the demodulation circuits. The first step in demodulation is to separate the subcarriers frequency-wise through the use of frequency-selective filters. The signals are then of the same form as before their linear combination in the modulation equipment, that is, a modulated subcarrier and a reference subcarrier at exactly half frequency. Doubling the reference subcarrier frequency produces an unmodulated signal of frequency exactly equal to that of the modulated subcarrier.

These signals are then compared in a conventional balanced phase detector. The differentially-connected outputs of the two halves of the balanced phase detector comprise the output signal of the channel.

Mode of Operation

Figure 2 indicates qualitatively the operation of the demodulation circuit. An arbitrary control signal to be transmitted is shown as the first curve. After the balanced modulation process this information is contained in the amplitude and phase polarity of the modulated subcarrier as shown in the second curve of Figure 2. The subcarrier amplitude is proportional to the control signal and the amplitude is zero for intervals of zero control signal. Further, there is a subcarrier phase reversal whenever the control signal polarity changes. The cross-hatching indicates the relative phase of the subcarrier. This signal is applied to one pair of input terminals of the balanced phase detector as shown.

The reference or unmodulated subcarrier is applied to the other pair of phase detector terminals. This subcarrier is of the same frequency as the modulated subcarrier, having been derived from the same oscillator. It has been transmitted at half modulated subcarrier frequency and has passed through a frequency doubler. Its phase relative to the modulated subcarrier phase is represented by the cross-hatching.

The envelope of E_{OA} is derived by adding the modulated signal to the reference subcarrier, taking into consideration the relative phases. Similarly, the envelope of E_{OB} is derived by subtracting the modulated signal from the reference subcarrier, again taking into consideration the relative phases. E_{OA} and E_{OB} are readily recognized as ordinary double sideband amplitude modulated signals; the only difference between the two is that the phase of the envelope of the one is inverted with respect to that of the other.

The voltages E_{OD} and E_{OC} represent the outputs of the two diodes, assuming peak detection. E_{OD} and E_{OC} obviously follow the envelopes of E_{OA} and E_{OB}, respectively. The DC level due to the reference subcarrier is the same on both sides of the circuit; the modulation phases are the inverse of each other. Forming the difference $E_{OD} - E_{OC}$ removes the effect of the reference subcarrier from the system output. The output is a replica of the control signal in both amplitude and polarity. It is apparent that the output zero is obtained independently of the level of the reference subcarrier. Therefore the system zero stability is, within practical limits, independent of the system amplitude or gain stability. This is to be compared with a frequency or phase modulation system where the zero stability is a direct function of the system frequency or phase stability.

Phase Detector Analysis

In order better to illustrate how the zero stability of the system comes about, Figure 3 shows the phase detector circuit again, this time with a few pertinent equations. This is probably the more or less standard phase detector analysis. The modulated subcarrier is applied to the two diodes in push-pull and the reference subcarrier is inserted in parallel. The phase angle ϕ represents the relative phase between the modulated and reference subcarriers. In a phase system this angle would itself be the modulation variable. The amplitude e_s of the modulated subcarrier is, in the present case, the modulation component to be detected. The angle ϕ is carried along in order to evaluate its effect on the demodulation process, the circuit being, after all, a phase detector.

The first pair of equations in Figure 3 gives the two voltages E_{OA} and E_{OB}, which are the signals applied to the two diodes. The second pair of equations is the result of simple trigonometric manipulation of the first. The angles α and β are functions of the angle ϕ and the ratio e_s/E_c and are meaningless here since the diodes are sensitive only to the amplitudes of the applied signals. The next pair of equations gives the diode load voltages E_{OD} and E_{OC}, assuming peak

detection. The expressions have been derived from the previous pair by using the amplitude terms only and simplifying by performing the indicated squaring operations.

The output of the circuit is obtained by subtracting E_{OC} from E_{OD}. The last equation represents the result, normalized with respect to the factor $2E_c$.

In a phase modulation system, the same phase detector circuit might conceivably be used. In that case both e_s and E_c would be maintained constant by amplitude limiting, and their ratio would therefore be constant. For zero modulation the arrangement of preceding circuitry would be such that the angle ϕ would be 90°, resulting in an output null. For the AM system under discussion the circuit arrangement would set and maintain ϕ at zero nominally. The output null is obviously obtained for e_s equal to zero.

Computed Response

The curves plotted in Figure 4 are computed from the last equation of Figure 3. These curves indicate the effect of the phase angle ϕ on the phase detector output when the circuit is used as an amplitude detector. It is seen that for ϕ equal to zero the output varies linearly as the amplitude of the modulated subcarrier. As ϕ departs from zero the output departs from linearity to a degree depending upon how far ϕ drifts. For ϕ equal to 90° there is no output for any e_s/E_c, and for ϕ greater than 90° the output polarity changes. The output for negative e_s/E_c is not shown but is negative. The complete response curve is skew-symmetric about the origin.

For any value of ϕ the response curve passes through the origin. This indicates that there are no stringent requirements on the phase stability of the system and, in particular, on the subcarrier filters, insofar as zero protection is concerned.

Effect of Phase Shift

It is evident that the primary effects of incidental phase shift are a departure from linearity and a change in gain. The effect of phase shift on system gain might be evaluated by computing the derivative of the response at the origin and using this derivative to define a gain. The absolute deviation from linearity may be defined as the percentage departure of the response from this linear response. Evaluating this derivative discloses that the incremental gain at the origin is proportional to cos ϕ. For an absolutely linear response, the incremental gain would be proportional to cos ϕ for all values of e_s/E_c up to unity. The curves of Figure 4 have shown that the response is not linear for ϕ unequal to zero. The curves of Figure 5 indicate the departure from absolute linearity, or in effect the amount of curvature of the response for various values of ϕ. Thus, for a fixed 10° phase error in insertion of the reference subcarrier, the maximum departure from linearity is about six per cent. If the maximum modulation depth is limited to 0.2 only about one per cent departure from linearity occurs for a fixed phase error of 45°.

These curves are somewhat misleading since they directly indicate only the effect of a fixed error in the insertion angle of the reference subcarrier. A more practical problem is to determine what happens if, after the system gain has been set up with ϕ equal to zero, ϕ subsequently drifts to some other value. Figure 6 shows what can be expected under these conditions. These curves are plotted on the basis of a percentage departure from the ideal response obtained for ϕ equal to zero. From the curve for ϕ equal to 10°, a maximum departure of four per cent from the ideal response will result if the modulation depth e_s/E_c is limited to values less than 0.8. If the modulation is limited to 0.2, a phase drift of 20° will result in about a six per cent departure from ideal.

This family of curves does not pass through the origin, although the response itself does pass through the origin. The intercept is $100(1 - \cos \phi)$. The fact that a finite percentage departure from zero is still zero allows the response to pass through the origin.

Limitations

It has been pointed out that reasonable linearity and good zero protection are obtained without excessively stringent requirements on the phase or frequency stability of the system. Naturally, this does not say that in the practical system there are no problems. The system zero drift is due principally to the zero drift of the modulation and demodulation circuits. Similar circuits might be used in a phase system but then the system phase drifts would also contribute directly to the zero drift. In the amplitude modulation system a small zero offset occurs as the result of the generation of the second harmonic of the reference subcarrier which is transmitted at half frequency. This offset is inherently constant for periods of time short as compared with component life and consequently does not contribute appreciably to zero drift. However, any zero balancing adjustment required in the output of the phase detector to calibrate out distortion or unsuppressed carrier or the equivalent causes the phase detector itself to be operated in a slightly unbalanced condition. Zero output then does not occur at the natural null of the phase detector. As a result phase drifts in the system do contribute slightly to zero drift.

The DC to DC gain can be no more stable than the AC gain of the entire system including the radio link. An FM radio link with voltage regulated audio and modulation circuitry in the transmitter and saturated, voltage regulated limiters in the receiver probably gives the best amplitude stability obtainable in that portion of

the system. Careful use of voltage regulation and temperature compensation of negative feedback amplifiers are the best approach to AC gain stability in the subcarrier equipment. The AM system is probably at a slight disadvantage where extreme gain stability is considered more important than zero stability.

Experimental Results

A number of experimental models of subcarrier equipment for proportional data transmission have been built using the AM system. These equipments have been laboratory tested under a broad range of service conditions. The zero drift is less than ±0.01 per cent of full scale signal per minute after ten minutes of warm up operation. The zero resolution is better than ±0.5 per cent of full scale signal. The dynamic response is flat within ±1.5 db to 30 cps, limited primarily by phase detector output filtering in present models.

The system linearity is defined in such a manner as to lump absolute linearity with gain stability. Figure 7 illustrates the definition. The horizontal axis represents the system input signal and is marked in per cent of full scale signal. The vertical axis represents the departure of the system output from the ideal linear response. The vertical scale is marked in per cent of full scale output. The inclined lines represent limiting values of the response for all variations of service conditions. A typical response is shown.

An ideal system would have no zero drift, the response would be absolutely linear, and the gain would be constant. This ideal response would be represented by a line coincident with the horizontal axis. A departure from linearity results in curvature of the response line, a zero drift shows up as a vertical shift in the response, and a gain change rotates the entire response curve approximately about the origin.

Conclusion

The elements of a simple proportional data transmission system using subcarrier amplitude modulation have been briefly discussed. Modulation and demodulation are relatively easily accomplished using well-known circuitry. For systems where zero protection is considered more important than extremely good linearity or gain stability, the AM system has many advantages.

Acknowledgment

Acknowledgment is due D. W. Ford for his assistance in the development of the experimental models.

Reference

Greenwood, Holdam, and MacRae, Electronic Instruments, MIT Rad. Lab. Series , Vol. 21, McGraw-Hill Book Company, Inc., pp 378-386, 1948.

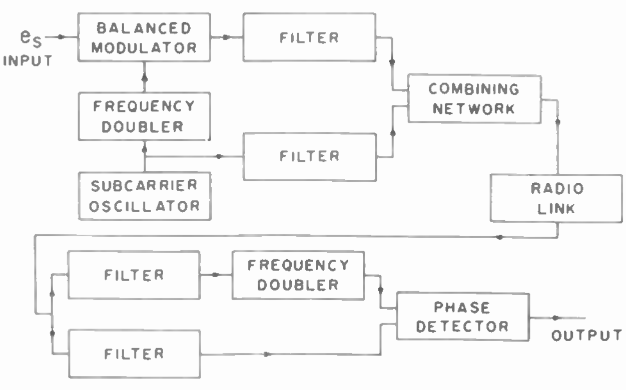


Fig. 1 - Simplified block diagram.

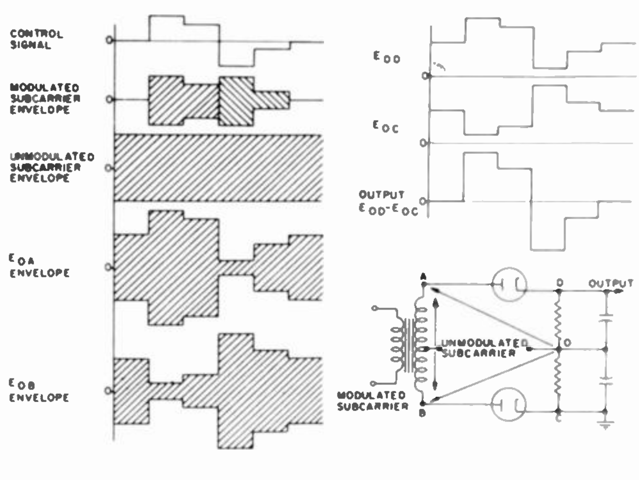


Fig. 2 - Phase detector - mode of operation.

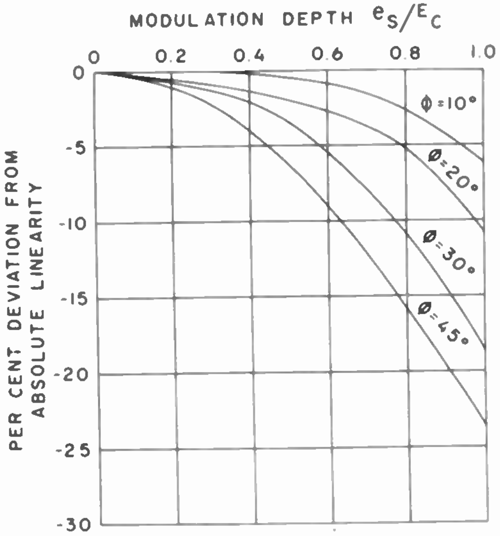


Fig. 5 - Departure from absolute linearity.

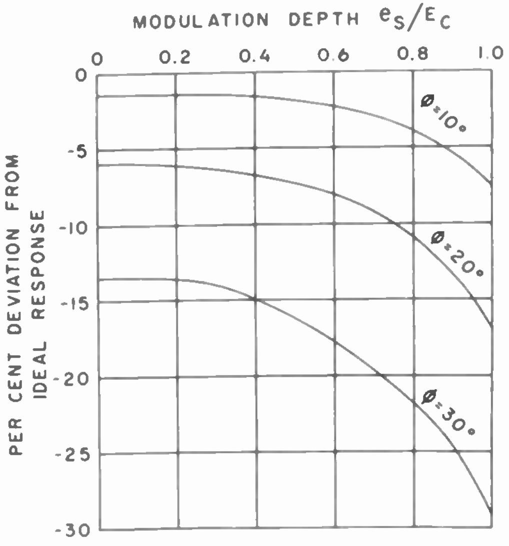


Fig. 6 - Departure from ideal response.

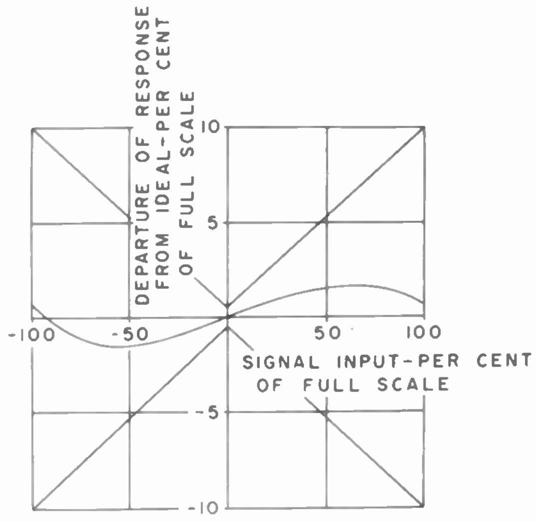
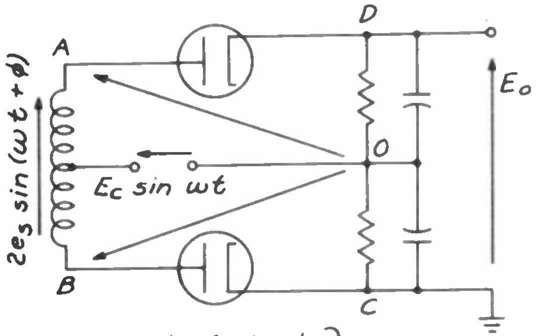


Fig. 7 - System performance.



$$\begin{cases} E_{OA} = E_c \sin \omega t + e_s \sin(\omega t + \phi) \\ E_{OB} = E_c \sin \omega t - e_s \sin(\omega t + \phi) \end{cases}$$

$$\begin{cases} E_{OA} = E_c\left[\left(\cos \phi + \frac{e_s}{E_c}\right)^2 + \sin^2 \phi\right]^{\frac{1}{2}} \sin(\omega t + \alpha) \\ E_{OB} = E_c\left[\left(\cos \phi - \frac{e_s}{E_c}\right)^2 + \sin^2 \phi\right]^{\frac{1}{2}} \sin(\omega t + \beta) \end{cases}$$

For Peak Detection

$$\begin{cases} E_{OD} = E_c\left[1 + \left(\frac{e_s}{E_c}\right)^2 + \frac{2e_s}{E_c} \cos \phi\right]^{\frac{1}{2}} \\ E_{OC} = E_c\left[1 + \left(\frac{e_s}{E_c}\right)^2 - \frac{2e_s}{E_c} \cos \phi\right]^{\frac{1}{2}} \end{cases}$$

$$\frac{E_0}{2E_c} = \frac{1}{2}\left\{\left[1 + \left(\frac{e_s}{E_c}\right)^2 + 2\frac{e_s}{E_c} \cos \phi\right]^{\frac{1}{2}} - \left[1 + \left(\frac{e_s}{E_c}\right)^2 - 2\frac{e_s}{E_c} \cos \phi\right]^{\frac{1}{2}}\right\}$$

Fig. 3 - Phase detector analysis.

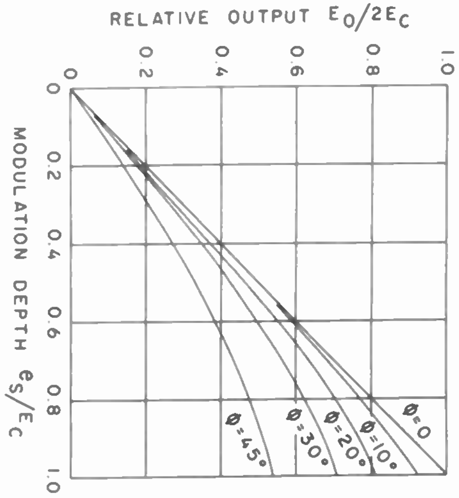


Fig. 4 - Computed response.

A DIGITAL AUTOPILOT COUPLER

W. L. Exner and A. D. Scarbrough
Hughes Research and Development Laboratories
Culver City, California

Summary

A system has been developed which automatically controls the heading of an aircraft in response to the output of a digital computer. This involves the use of a coupler unit which converts the binary steering signal supplied by the computer to a form acceptable to a conventional autopilot.

Factors which must be given consideration in the design of such a coupler include the operational features of the computer and autopilot, and the potential instability due to the finite iteration time of the computer. A successful coupler is described and its performance discussed.

Introduction

The autopilot coupler which will be discussed in this paper was developed as a part of an automatic, precision, aerial navigation and weapon control system. This system, which was developed and successfully flight tested under a U.S.A.F. contract, is known as Digitac. One of the noteworthy features of this system is that it marks the first successful use of an airborne digital computer to control an aircraft.

This paper will deal with some of the problems associated with the use of a digital computer in such a control system, and describe the autopilot coupler which was developed for this application. Two other papers, describing the Digitac computer and the flight testing of the complete Digitac system, are being presented at this convention.[1,2]

Problems of Digital Control

In contemplating the use of a digital computer in any automatic control system careful consideration must be given to two characteristics of digital computation; namely, quantization and delay. Since the computer requires a finite time to solve the control problem, it can provide new values of the control function only once each computation cycle, and the output is therefore quantized both in time and in magnitude. Fluctuations in the output signal caused by discontinuities of the input data can be minimized by proper programing of the computer, but spurious transients resulting from computer malfunction can best be rejected by analog filtering of the output signal. Although such errors should be infrequent and of short duration, the fact that the computer is being used to control an aircraft makes it imperative to include sufficient analog filtering to protect the aircraft, and such filtering results in increased delay.

The delay inherent in digital control systems is a result of the time required to compute the control function, the time required to smooth the data, and the time required to filter the output signal.

The time required to smooth the data and compute the control function depends on the speed of the computer, the quality of the input data, the complexity of the problem and the skill with which it has been coded. Although the Digitac computer can add 2640 sixteen-digit binary numbers per second and perform other arithmetic operations with corresponding rapidity, the complexity of the problem is such that its solution requires about one-half second. Thus it is apparent that if the computer is to perform this control function satisfactorily it must be required to pass only those frequencies substantially below one cycle per second.

Problems of Aircraft Control

The basic aircraft-autopilot control loop is shown in Figure 1. Here the heading of the aircraft is measured by the compass, which supplies the proper control signal to the autopilot to enable it to hold the heading constant. If a change in heading is required, it must be made manually through the autopilot turn control, after which the system will hold the aircraft on the new heading.

Perhaps the most obvious way to incorporate a navigation computer into such a system is shown in Figure 2. Here the computer receives input data consisting of heading, air speed, altitude, and ground position, from which it computes the required heading and, through a digital-to-analog converter and filter, supplies the autopilot with a steering signal to achieve this heading.

The basic function of the computer is to compute a desired heading which, since it depends on the aircraft position relative to its destination, and the drift due to wind, is a slowly varying function of time. Since in the system shown in Figure 2 the computer and filter are included in the aircraft-autopilot stability loop, the computer is required to pass all frequencies handled by this loop, some of which are many times higher than those associated with changes in the desired heading. This system therefore puts much more stringent requirements on the digital computer than its basic function demands.

In the Digitac system, the stability problem has been approached from a different point of view, as shown in Figure 3. Here the burden of providing aircraft stability is carried by the unbroken analog control loop consisting of aircraft, compass, and autopilot; and the computed heading error is added as a relatively slowly varying quantity. Although the computer output is heavily filtered to minimize the effect of spurious transients, this filtering which is outside the control loop, has no effect on the aircraft-autopilot stability. It is apparent that the use of this method tends to simplify both computer and filter design, since it now becomes advantageous to filter out high frequency components of the computed steering signal which might be generated in turbulent air. Furthermore, digital prediction and smoothing techniques can be employed with greater success when applied within this limited bandwidth.

It is fundamental to the operation of this system that the rate at which variations in steering signals are transmitted through the autopilot coupler be slow enough to provide adequate filtering, but fast enough to follow normal course changes. When radical course changes are required, the coupler saturates and holds the aircraft at a fixed bank angle until its heading approaches the heading required. This will be discussed further, after a description of the coupler equipment.

The Digitac Coupler

The E-6 autopilot, to which the Digitac coupler was tailored, is a precision autopilot designed primarily for the control of bombardment aircraft. It is a stable and precise autopilot and has excellent dynamic characteristics. All heading control is achieved through coordinated turns, the bank angle being a direct function of the heading error, up to a predetermined maximum bank, which for the Digitac tests was set at 20 degrees.

The heading reference for the autopilot was supplied by a J-2 Gyrosyn compass, the transmission between the compass and autopilot being by means of a pair of synchros. In order to modify this heading reference in response to the steering signal, it was only necessary to interpose a differential synchro between the compass and the autopilot and arrange for the shaft position of this differential to be positioned in accordance with the steering signal. The heading reference, as seen by the autopilot, is thus the compass heading plus-or-minus an angle which is a quasi-integral of the computed steering signal.

In addition to the differential synchro, the Digitac autopilot coupler consists of equipment and circuitry which smooths the digital steering signal supplied by the computer and positions the shaft of the differential synchro in response to the resultant function.

A block diagram of the complete coupler is shown in Figure 4. The mechanical drive for the differential synchro is provided by a bi-directional notching motor, whose shaft revolves 1/50 of a revolution for each applied voltage pulse, and is locked in position at all other times. Direction of rotation is determined by a reversing relay, which is controlled by the sign of the computed steering signal. Since the notching motor is coupled to the differential synchro through a 36:1 gear train, each pulse applied to the motor corresponds to a heading increment of 1/5 of a degree.

Application of pulses to the notching motor is controlled by a seven-digit-plus-sign shifting-counting register, as shown in Figure 4. The gate circuit allows pulses from the pulse generator to be applied simultaneously to the "count" input of the register and to the notching motor, except when the contents of the register are zero. Since the register is designed to count down, a number shifted into it (representing the steering signal) is then counted down to zero, whereupon the pulses are stopped. During this counting operation the notching motor and the differential synchro will have been rotated through an angle proportional to the magnitude of the steering signal. It should be noted, however, that except for very small steering signals, there is never time for a complete count-down to occur. A new steering signal, representing the latest computed information, is shifted into the register each half-second, regardless of whether the previous count-down was completed or not. Thus, spurious steering signals of high amplitude but short duration have little effect on the autopilot.

In order to achieve the rapid correction of heading errors up to the full capacity of the register, without danger of over control when the steering signal approaches zero, it has been found desirable to make the pulse rate proportional, within limits, to the steering signal. Since a voltage proportional to the steering signal is developed elsewhere in the system for operation of the PDI (Pilot Director Indicator), this voltage is used to control the repetition rate of the pulse generator shown in Figure 4, over a range of approximately four to fifteen cycles per second.

Figure 5 is a photograph of the two chassis which include most of the circuitry of the research model of the autopilot coupler. These chassis are plug-in units which fit into a rack along with the rest of the input and output equipment of the Digitac system. The upper chassis holds the shifting-counting register, and the lower chassis includes synchronizing and control circuitry. Figure 6 is a photograph of the

differential synchro and notching motor assembly, together with the circuitry required for the operation of the notching motor.

The coupler operating as described above has been demonstrated to be highly satisfactory for controlling the aircraft when the required heading changes are within the ± 12.6 degree capacity of the shifting-counting register. However, the Digitac system must be able to control the aircraft over a pre-programed course, involving heading changes of up to 180 degrees, and for such heading changes a different mode of operation is required.

In order to explain how this is accomplished, it will be necessary to describe the shifting-counting register in somewhat greater detail. Although the register consists of seven stages, only six of these are required to accommodate its full capacity of ± 12.6 degrees, each bit being equal to 1/5 of a degree. The seventh or most significant stage, contains a "one" whenever the computed steering signal indicates a heading error greater than this saturation value. Since there is no count propagation into the overload stage, whether it contains a "one" or a "zero" is determined solely by whether the last computed steering signal is greater or less than saturation. Whenever such saturation does occur, the pulse gate is held open and simultaneously the repetition rate of the pulse generator is increased to approximately 30 per second. This causes the differential synchro to be rotated faster than the maximum rate of turn allowed by the autopilot, with the result that the aircraft is held in a fixed 20 degree bank.

During the turn a new steering signal is shifted into the register every half-second. As the turn progresses, the magnitude of these numbers progressively decreases, until eventually one appears which is within the normal capacity of the register. This indicates that the aircraft is within approximately 12 degrees of the required heading. As soon as this occurs the pulse gate is closed for five seconds, stopping the notching motor and allowing the aircraft to roll out of its bank. In accomplishing roll-out, the aircraft turns through an additional 12 degrees which brings it to approximately the required heading; so that when, at the end of the five second delay, the coupler returns to its normal mode of operation only very small heading corrections remain to be made.

Conclusions

When digital computers are applied to automatic aircraft guidance systems, their limited bandwidth may be insufficient to achieve stable operation if the computer is required to handle the entire control problem. The most promising approach to the stability problem appears to be to recognize that it consists of two nearly independent parts; namely, the aircraft stability problem, and the guidance stability problem.

The aircraft stability problem involves a wide band of frequencies, but the computations are simple and of a type which can readily be solved by analog techniques, as exemplified in conventional autopilots. The guidance stability problem is much more complex and, since several modes of operation may be required, the flexibility of digital computation is desirable. Fortunately, the required bandwidth is well within the capabilities of moderate speed digital computers.

The method proposed in this paper for combining the two types of computation has been proven by extensive flight tests to provide smooth and accurate control, even during periods of operation when for one reason or another the steering signal was so erratic as to make it almost impossible for the pilot to hold the aircraft on course by following the PDI. Once the optimum range of pulse rates had been determined there was no tendency for the system to over-control, yet required heading corrections were accomplished by the coupler at least as efficiently as by the human pilot.

The design of an autopilot coupler for use with any particular system must inevitably depend to a considerable degree on the peculiarities of the computer and the autopilot with which it will be used. However, it is believed that the basic principles of the Digitac coupler will prove widely applicable, not only to other airborne digital control systems, but to many other automatic control systems where it is desired to take advantage of the accuracy and flexibility of digital computation.

---

[1] E. E. Bolles, "The Digitac Airborne Digital Computer"

[2] E. M. Grabbe, D. W. Burbeck, and S. B. Neister, "Flight Testing of an Airborne Digital Computer"

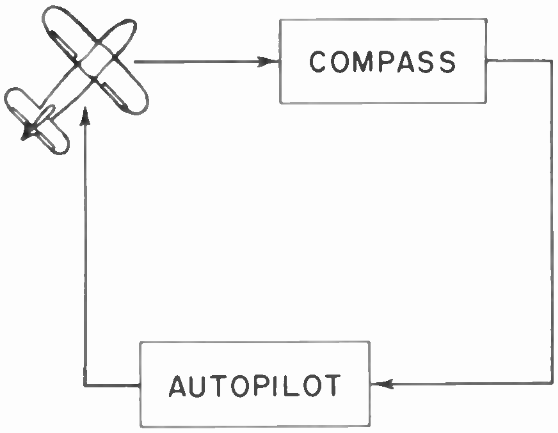


Fig. 1 - Basic heading control loop.

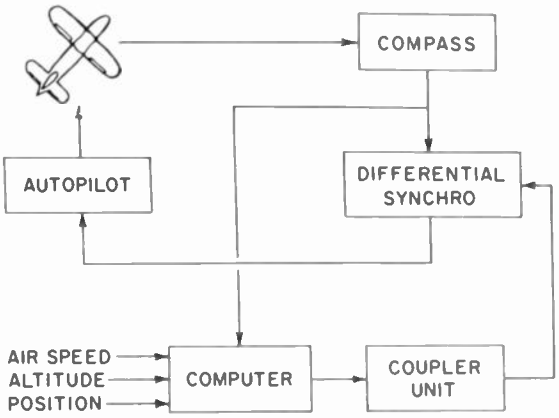


Fig. 2
Control loop, including navigation computer.

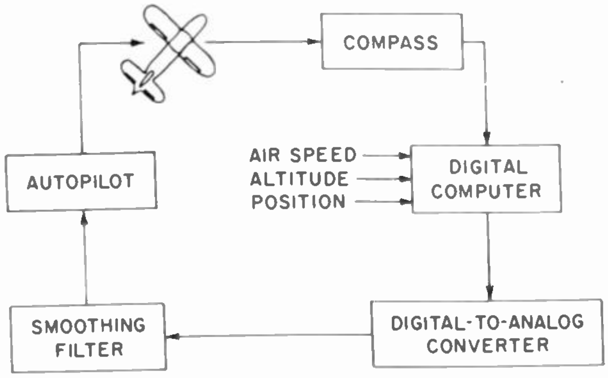


Fig. 3 - Digitac heading control.

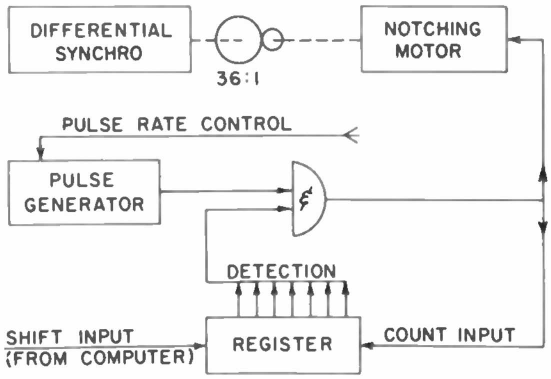


Fig. 4 - Digitac coupler block diagram.

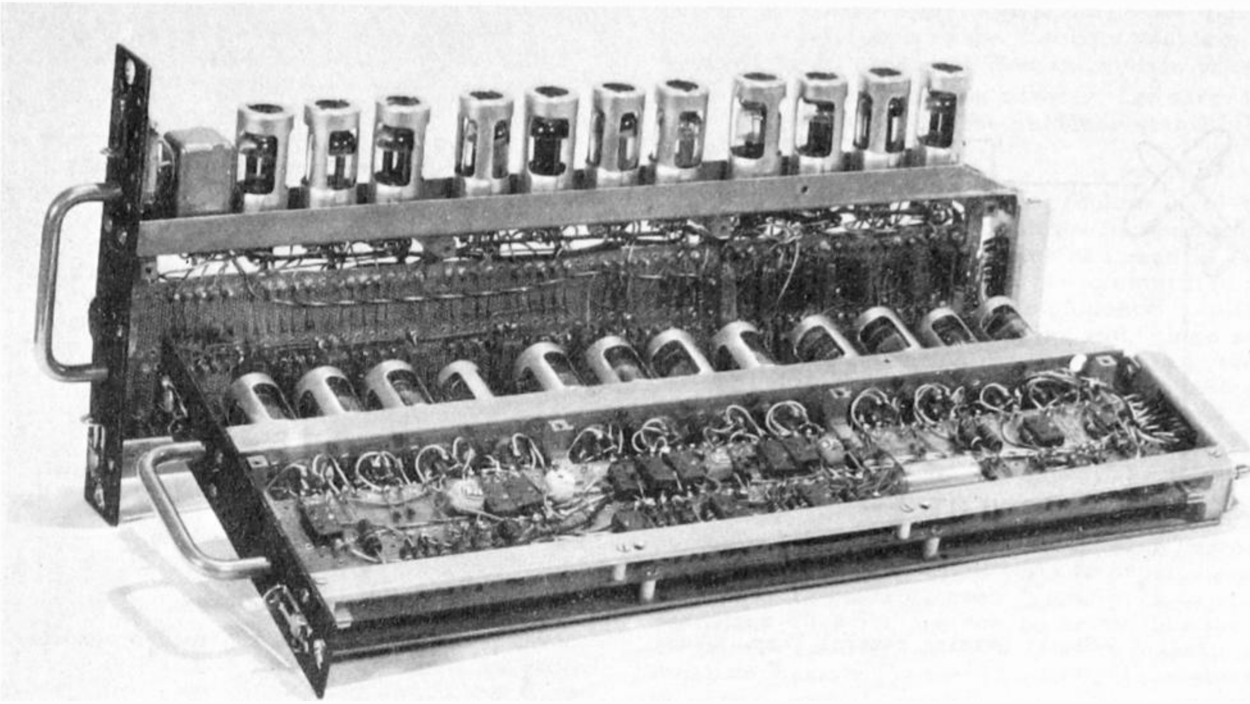


Fig. 5 - Register and control chassis.

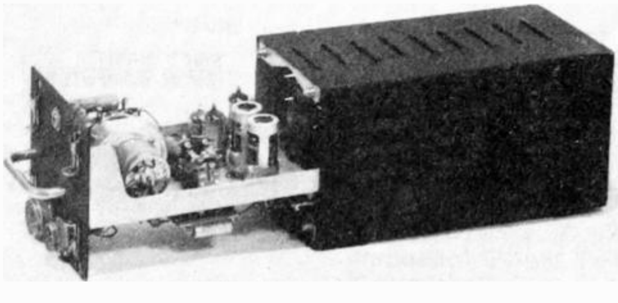


Fig. 6 - Differential-synchro drive assembly.

SYSTEM COMPENSATION WITH A
DIGITAL COMPUTER

John M. Salzer
Hughes Research and Development Laboratories
Culver City, California

Introduction

Recent years have witnessed a phenomenal growth of the digital computer field. The usefulness of these high-speed computers in scientific studies and in business systems is well established, while their use in automatic control systems is relatively unexplored. There is little question, however, that digital techniques are making their way into the control field and that studies directed toward a better understanding of the dynamic interaction between digital and analog elements are justified.

This paper is a report on certain theoretical and experimental studies conducted in regard to compensating a simple closed-loop system by appropriate programs in the digital computer, which was one of the elements in the system. It is assumed that the use of a digital computer has been decided upon for various reasons, such as the data transmitted being already in digital (pulse-coded) form, or the flexibility afforded by the use of a general-purpose machine, or the precision required in the calculations; the possibility of system compensation with the digital computer is then an additional advantage. The present study is limited to a particular system configuration and to the use of linear compensating programs.

The analytical studies regarding the systems and their compensations were tested experimentally on the laboratory setup shown in Figure 1. The analog portions of the system were simulated with the GEDA (Goodyear Aircraft) analog computer shown on the left. The results were recorded on the Midcentury recorder shown on the right. The digital computer and its associated test equipment is shown in the center of the figure. The digital computer used was the second laboratory model (Model II) built for a particular control system under development. It is a general-purpose serial binary machine with a word length of 17 binary digits, using a magnetic drum and operating at about 160 kilocycles per second pulse rate. The input-output equipment[1] of this model handles 9 different inputs and 4 different outputs. With a conversion performed on each drum revolution the complete input-output cycle (including one drum revolution for calibration) takes 14 drum revolutions or 0.105 seconds. In the experiments only one input and one output were used, as the computer was a single-channel link.

Analytical Representation of System

In all cases studied the controlled element, $G_2(s)$, consisted of an integration in tandem with a simple lag:

$$G_2(s) = \frac{K}{s(\alpha s+1)} \quad . \qquad (1)$$

Figure 2 shows how the digital computer was inserted into the closed-loop system, with means provided to shunt the computer for a ready check against the well-known behavior of the all-analog loop.

An important effect of the digital computer on the data passing through it is the sampling of the data. Due to the fact that several operations are performed in sequence on the data entering the digital computer, it is clear that the computer accepts information only intermittently. The minimum period between samplings is the computation time. The computation also introduces a delay between the time an input sample is accepted and the time the corresponding output sample is obtained. This output sample is then clamped in order to supply a continuous signal to the controller. The equivalent representation of the digital computer is, therefore, that shown in Figure 3.

The representation of sampling by impulse modulation, where the carrier, $u^*(t)$, is a series of impulses occurring T seconds apart, has been discussed in the literature.[2,3] Accordingly, the transforms $F^*(s)$ and $F(s)$ are related by

$$F^*(s) = \frac{1}{T}\sum_{k=-\infty}^{\infty} F(s+jk\Omega) \qquad (2)$$

where $\Omega = 2\pi/T$ = sampling angular frequency. The effect of sampling is the creation of an infinite number of side bands around Ω and all harmonics of Ω. To regain the continuous function these extraneous (complementary) frequencies need to be eliminated. The clamping unit at the output of the digital computer serves this purpose, although it performs only an imperfect job of low-pass filtering. The unit impulse response of clamping is a rectangular pulse lasting the sampling interval T. Such a pulse can be expressed as the difference between a unit step at $t=0$ and unit step at $t=T$, so that the transfer function of the clamping unit is $(1/s) - (e^{-Ts}/s)$.

Since the transfer function of the time delay of γT seconds is merely $e^{-\gamma Ts}$, the analytical model of the digital computer is an impulse modulator followed by a box (of suitable color) having the transfer function

$$G_1(s) = \frac{e^{-\gamma Ts}(1-e^{-Ts})}{s} \quad . \qquad (3)$$

The only thing not yet accounted for in the system is the compensating program itself. We will show subsequently that in the linear case such a program can also be represented by a transfer function. If so, everything in this closed-loop system can be characterized by a transfer function, except the sampling process itself. The net result of this is that the forward loop consists of an impulse modulator followed by some over-all transfer function G(s) which includes the compensating program, the computational delay, the clamping filter, and the controller.

In references 2 and 3 it was shown that the relation between the controlled quantity, C(s), and the sampled form of the reference quantity, $R^*(s)$, is

$$\frac{C(s)}{R^*(s)} = \frac{G(s)}{1+G^*(s)} \qquad (4)$$

for the conditions just discussed. The starred quantities are related to the unstarred ones according to Equation (2). It is often sufficient to find the sampled output in terms of the sampled input, for which the relationship is

$$\frac{C^*(s)}{R^*(s)} = \frac{G^*(s)}{1 + G^*(s)} \quad . \qquad (5)$$

The effect of sampling, the time delay due to computation, the lag and imperfectness of the clamping filter are disadvantages of using this digital computer. Depending on the application these effects may or may not be important, or they may be accepted in exchange of the advantages gained by the use of a digital computer. In the present investigations the parameters of the system were purposefully so chosen as to make the detrimental effect of the computer quite apparent; the compensating potentialities of the computer could then be demonstrated quite strikingly.

Transfer Function of Computer Program

In conventional servo systems consisting of a single loop, compensation is accomplished by the insertion of some kind of filter (analog filter) in the forward or feedback path or both. The transfer function W(s) of such a filter of the linear kind is a rational function of s. Another way of describing such a filter is by its response w(t) to a unit impulse at t = 0, where w(t) is the inverse transform of W(s).

Knowing w(t) of a particular filter certainly allows one to find the output of this filter if the input consists of a sum of impulses. However, the input i(t) is often continuous, and in order to obtain the continuous output o(t) it is necessary to replace i(t) by a set of impulses, as indicated in Figure 4. Let these impulses be spaced T seconds apart and assign to them values of areas equal to the approximate area under i(t) between this impulse and the previous one. For example, the impulse at $t = \ell T$ has the value $T.i(\ell T)$ which is the rectangular area cross-hatched in Figure 4. The value of the output at $t = \ell T$ will be the sum of the responses of the filter to the individual input impulses; that is,

$$o(\ell T) = \sum_{k=0}^{\infty} T.i\left[(\ell-k)T\right].w(kT) \qquad (6)$$

The correct value of $o(\ell T)$ is approximated better as T in the above equation is made smaller. The expression for the output will be correct when T goes to zero. As T goes to zero, ℓ and k approach infinity for the same interval in such a manner that the products ℓT and kT stay constants, say t and τ. In the limit Equation (6) becomes

$$o(t) = \int_0^{\infty} d\tau . i(t-\tau).w(\tau) \quad , \qquad (7)$$

which is recognized as the famous superposition integral.

The purpose of this derivation was not to obtain the superposition integral, which is derived in many available books, but to impress on the listener the fact that in obtaining the output of a continuous-data filter he had to resort to the subterfuge of input impulses, impulse response and summation process. In characterizing the action of a digital computer program, one can save a step in the above derivation and use Equation (6) itself, because the input of the program is a set of impulses (samples) and the output another set of impulses. The use of impulses now is not a subterfuge but the actual situation. The sampling interval T stays finite and the effect of the program is described by the weighting sequence w(kT) or w_k, rather than a weighting function $w(\tau)$.

The embarrassing thing in Equation (6) is that the weighting sequence is infinite so that the computation of a new output sample would take "too long"; such a program is not realizable. In practice we can truncate the sequence, but we can do even better by adding terms which make use of previously computed output samples also. Thus, a completely general, yet realizable, linear program would correspond to the equation

$$o(\ell T) = \sum_{k=0}^{m} a_k i\left[(\ell-k)T\right] - \sum_{k=1}^{n} b_k o\left[(\ell-k)T\right] . \quad (8)$$

If the output of the program is thought of as a sequence of impulses, $o(\ell T)$ is the area under the impulse occurring at the output of the program at $t = \ell T$. One may just as well call this whole sequence of impulses $o^*(t)$, since it is a legitimate though ragged function of time, and similarly for the input. Doing so changes Equation (8) into

$$o^*(t) = \sum_{k=0}^{m} a_k i^*(t-kT) - \sum_{k=1}^{n} b_k o^*(t-kT) \quad . \quad (9)$$

The transform of the above equation is readily

obtained, if the transforms of $i^*(t)$ and $o^*(t)$ are called $I^*(s)$ and $O^*(s)$ respectively and if it is remembered that a time delay x is represented by multiplication of the transform by e^{-xs}. Thus,

$$O^*(s) = \sum_{k=0}^{m} a_k e^{-kTs} I^*(s) - \sum_{k=1}^{n} b_k e^{-kTs} O^*(s) , \quad (10)$$

from which the transfer function of the program is derived as

$$W^*(s) = \frac{O^*(s)}{I^*(s)} = \frac{a_0 + a_1 e^{-Ts} + \dots + a_m e^{-mTs}}{1 + b_1 e^{-Ts} + \dots + b_n e^{-nTs}} \quad (11)$$

In what follows the letter z will be used in place of e^{-Ts}; thus, Equation (11) can be written as

$$W^*(s) = \frac{a_0 + a_1 z + \dots + a_m z^m}{1 + b_1 z + \dots + b_n z^n} \quad (12)$$

Having obtained a transfer function for a linear, real-time digital-computer program, one may deal with it as he would with the transfer function of any filter. The frequency analysis of sampled-data filters (of which digital-computer programs are one example) has been discussed in the literature[4,5]. In what follows the application of these techniques to the digital compensation of the particular closed-loop systems studied will be discussed.

Particular Systems Studied

As noted before, the analog portion of the system had the transfer function $G_2(s)$ of Equation (1). Two sets of parameters α and K were chosen for the purpose of investigation, and the two systems were called narrow-band and wide-band, the latter having a nominal bandwidth four times that of the former. The two sets of parameters are listed:

Case	α sec	K	ω_n rad/sec	f_n cps	ζ
Narrow-band	0.42	4.2	π	0.5	0.38
Wide-band	0.056	8.9	4π	2.0	0.7

In one case the resonant frequency of the closed-loop all-analog system is seen to be 0.5 cps, while in the other case it is 2.0 cps.

Since the computation in the computer was synchronized with the input-output cycle, the sampling time used was T = 0.105 sec, which corresponds to a sampling frequency of $f_s = 1/T$ = 9.52 cps. Theoretically the bandwidth of the system should not be greater than $f_s/2$ = 4.76 cps, but in a practical servo system one would want to stay well below this band. From this point of view even the narrow-band case is critical, since the bandwidth of the system is greater than its resonant frequency. The wide-band system is even more difficult to handle.

The two cases will be discussed separately. After indicating the effect of inserting the digital computer in each system, the compensation of the mixed digital-analog system will be illustrated.

Compensation of the Narrow-Band System

The locus of the open-loop transfer function of a system (the Nyquist plot) is a very useful guide to estimating the general behavior of the system. In particular, a good picture of system stability can be rapidly obtained from these plots, which also point to the nature and degree of compensation required. For the narrow-band system the plots of Figure 5 show the effect of inserting the digital computer. The first curve is the open-loop locus of the all-analog system, $G_2(s)/K\alpha$; the other curves are the open-loop sampled loci of the mixed system, $G^*(s)/K\alpha$, with two different simulated computing delays, without any compensation. The deterioration of the system is due mainly to the lags introduced by the simulated computing delay and the clamping device. The sampling process itself has an almost negligible influence in this narrow-band system. Intersection of the curves with the negative real axis gives the limit $-/K\alpha$ for stability, as summarized:

Case	Delay	$-1/K\alpha$	K_{lim}
Analog	0	0	∞
Mixed	3 Drum Revolutions	-.175	13.6
Mixed	13 Drum Revolutions	-.35	6.8

The experiments verified these limits accurately.

The compensation of this system was not a difficult task. The mixed system with the shorter (3 drum revolutions) artificial delay was chosen as the object of the study. It is evident from the plots of Figure 5 that the introduction of some phase lead in the system should restore some or even more than the stability lost. A prediction program compensating for a time delay equal to a full sampling period T (or 14 drum revolutions) was thought of as approximately the right amount of lead, for the effective delay of the clamping device is T/2 (or 7 drum revolutions), and the simulated computing delay was 3 drum revolutions. Although 7 + 3 drum revolutions make only 10, it must be remembered that no program can perform ideal prediction and, therefore, the compensation will be for less than the 14 drum revolutions design figure.

It can be shown that a parabolic approximation through three points of i(t) produces a predicted point o(t), which would be equal to i(t + T) ideally. The formula of this prediction is:

$$o^*(t) = 3i^*(t) - 3i^*(t - T) + i^*(t - 2T), \quad (13)$$

which after transformation yields

$$W^*(z) = \frac{O^*(z)}{I^*(z)} = 3 - 3z + z^2 \quad . \quad (14)$$

The latter is the transfer function of parabolic prediction and a special case of Equation (12). The absence of a denominator in Equation (14) means that the computation of $o^*(t)$ is unaffected by previously computed output samples $o^*(t - kT)$.

It can be shown that this classical formula is never optimum in the least-square sense, and moreover, there is no positive assurance that a prediction, even if ideal, would be the best lead compensation for this system. The program obtained above was, therefore, used mainly as a starting point in the design. The stabilization was improved experimentally by varying the constants 3 and -3 of Equation (13) and keeping their magnitudes equal. Based on the observed step response of the closed-loop system the optimum seemed to be

$$W^*(z) = 2.5 - 2.5z + z^2 \quad . \quad (15)$$

The locus of this program and its effect on the system transfer function are shown in Figure 6. The plot of $W^*(z)$ is only shown at low frequencies, but its locus is seen to be similar to the locus of a lead network. The high frequencies are accentuated (for example, at half the sampling frequency, 4.76 cps, the gain is 6), and therefore the noise problem, if any, is not helped. This is of course, the expected effect of just straightforward prediction.

The step response curves corroborated the above frequency analysis. Actually, as was later realized, a factor of two slipped into the program, and as a consequence, not only the bad effect of the computer was erased but this was done despite an increase by a factor of two in the gain of the system. The step responses of the all-analog system with K = 4.2, of the mixed system without compensation and with K = 4.2, and of the mixed system with prediction compensation and K = 8.4 are compared in Figure 7.

The relevant stability limits are tabulated herewith:

System	Delay in Drum Revolutions	Compensation	K_{lim}	
			Theoretical	Experimental
Analog	0	None	∞	---
Mixed	3 D. R.	None	13.6	13.6
Mixed	3 D. R.	$2.5-2.5z+z^2$	28	33

The small discrepancy in the last two figures is probably due to the approximation used in the analysis.

Compensation of the Wide-Band System

The compensation of the narrow-band system was more or less straightforward, at any rate not difficult. Actually, not much time was spent on that system, and most of the investigations were concentrated on the wide-band system to be discussed. Before the insertion of the digital computer, this system had a resonant frequency of 2.0 cps, which in combination with the 0.7 damping factor corresponds to a nominal bandwidth of about 2.6 cps. This is an appreciable portion of 4.76 cps allowed by the sampling theorem. In addition, the task was purposefully made more difficult by simulating a long computing delay when the digital computer was inserted into the system.

The effect of inserting the digital computer into the wide-band system are shown in Figure 8. One curve is the normalized open-loop locus of the all-analog system, $G_2(s)/K\alpha$ (where $\alpha = 0.056$); the other two curves are for the mixed system* with a simulated computing delay of 13 drum revolutions, or almost a complete sampling period. Since the locus of the mixed system function crosses the real axis at $-1/K\alpha = -2.2$, the maximum permissible K is 8.2 for stability. Thus, at the value K = 8.9 for which the all-analog system had a damping ratio of 0.7, this mixed system is already unstable.

Compensation for this system must not only restore stability but must also recapture a reasonable damping ratio. The qualitative nature of the desired compensation will be discussed first. It is clear that for improved stability of the mixed system, curve (c) of Figure 8 needs to be dented around the negative real axis for improved stability. This is a perennial problem in servo system design, and the linear solutions perennially call for a compensating filter which has (a) a phase lead in the critical frequency range, (b) attenuation in this range, or (c) both. The unsophisticated design of the filter often starts with some kind of a plausible network configuration, whose parameters are then adjusted analytically and experimentally. The same approach will be used also in the present case, except that the filter in question will be a digital-computer program.

The general form of the program transfer function is Equation (11) or (12). One special case of this is when all b's are zero and only the input samples are weighted in the computation. There are some rigorous and elegant analytical methods available to treat this case, but it becomes quickly apparent that for attenuation in the critical band (around 1.1 cps) this special filter is very wasteful of parameters; i.e., m needs to

*The two curves, $G(s)/K\alpha$ and $G^*(s)/K\alpha$, differ only by the effect of the sampling process itself, which manifests itself at the higher frequencies. Both curves include the effects of the computing delay and clamping.

be large and the resulting computation lengthy. For phase lead (prediction) this special-case program is quite suitable, but the gain at the higher frequencies (past 1.1 cps) becomes quite intolerable in view in view of the shape of the complex-plane plot. Not only the noise problem would be aggravated, but stability will hardly be improved because the phase lead brings the amplified higher frequencies to the negative real axis.

These initial considerations indicate two things: first, (unlike the narrow-band case) phase lead by itself is insufficient compensation; second, the program transfer function should have a non-unity denominator (not all b's should be zero), which means that not only input samples but also previous output samples (previous computed results) should be weighted in the computation. The next thing to decide is the complexity of the program; that is, the value of n. Here is where "experience" is needed for an adequate choice. The problem was to get considerable attenuation around 1.2 cps without much of a phase lag at this frequency. The phase lag must be concentrated at frequencies lower than 1.2 cps, and a phase lead (coupled with attenuation) is desired at 1.2 cps and above. It can be shown that n must be 2 or greater for this purpose. Further analysis showed that with n = 2 the compensation was inadequate, and n = 3 was chosen for the complexity of the program.

Figure 9 indicates the steps made in obtaining a reasonable starting point in the design of the compensating program. The first three steps develop a suitable denominator locus, while the fourth step is a reciprocation. These loci are readily sketched by noting that $e^{-jkT\omega}$ is a (counter-clockwise) unit circle and that the speed with which the circle is traversed is proportional to k. In part (a) only one exponential term is present so that the locus is truly a circle, only half of which is plotted corresponding to the range of 0 to $\Omega/2$ rad/sec, where $\Omega = 2\pi/T$ is half the sampling angular frequency (4.76 cps in this case). In part (b) of this figure a third-degree term is added. The vector construction corresponding to the point $\omega = \Omega/12$ or $\omega T = 30°$ (or 0.8 cps) is shown for illustration of the sketching technique. The effect of the third-degree term is to swing the locus away from the origin more rapidly at low frequencies; upon reciprocation this will give a narrower pass band. To exaggerate this effect a second-degree term is also added in part (c). This has two additional effects: it bunches the higher frequencies, making the attenuation more uniform in this range; it introduces a moderate phase lag (which will become a lead upon reciprocation) at higher frequencies. The reciprocation of (c) and suitable scaling give (d), which is then the locus of the program forming the starting point in our analog-experimental design:

$$W_o^*(z) = \frac{0.1}{1 - 0.3z - 0.3z^2 - 0.3z^3} \qquad (16)$$

As Figure 9(d) indicates, this program would achieve compensation mainly by attenuation. It will be demonstrated below that an improved version of this program leans more heavily on phase-lead type compensation.

When the program (16) was inserted into the computer, the system performance improved decidedly. However, one aim of the study was to develop an experimental method of improving the compensation. This could be done by varying the parameters of $W_o^*(z)$, which are numbers in the computer, but it would be difficult to tell directly how changes in the various coefficients of (16) would influence the behavior of $W_o^*(z)$. The steps of adjustment could not clearly defined. defined. On the other hand, if $W_o^*(z)$ is developed into its partial fractions, as

$$W_o^*(z) = \frac{0.049}{1-0.95z} + \frac{0.051\ (1 + 0.32z)}{1 + 0.65z + 0.32z^2}\ , \qquad (17)$$

the parameters are closely related to the poles and residues of the program and a logical sequence of adjustments is more readily evolved. It is possible to program $W_o^*(z)$ in such a manner that the coefficients of (17) appear explicitly as the numeric operands in the computer.

The loci of these partial fractions are shown in Figure 10. It is immediately clear that the first of the partial fractions of (17) determines the nature of compensation in a predominant manner. Nevertheless, the effect of the second fraction is important, because it counteracts the phase lag of the first one without impairing the attenuation significantly. With this picture in mind the succession of parameter adjustments were readily made, using the step response of the closed-loop system (with K = 8.2) as a criterion. After about half a dozen changes

$$W_1^*(z) = \frac{0.075}{1-0.895z} + \frac{0.45(1 + 0.33z)}{1 + 0.7z + 0.4z^2} = \frac{0.525 - 0.202z - 0.103z^2}{1 - 0.195z - 0.227z^2 - 0.358z^3} \qquad (18)$$

was obtained as the improved compensating program. The locus of this program as well as its effect on the open-loop system function are shown in Figure 11. The improved program, seen in part (b), shows that the experimental changes increased the role of phase lead with respect to attenuation.

Figure 12 shows the step responses of this system under the various conditions. The dotted line is the response of the all-analog system with K = 8.89. The dashed line illustrates the instability brought about by the insertion of the digital computer when the gain constant was held at K = 8.2. The solid line shows the effect of compensation with the improved program (18). Although the original performance has not been recaptured, compensation has stabilized the system and produced a

reasonable response. Furthermore, the computation for this compensation takes only a portion of the approximately 0.1 seconds allowed for computational delay. Thus, time is left for other calculations, which are assumed to have been the justification for using the digital computer in the first place.

The stability limits for the various conditions are summarized in the table below.

System	Delay in Drum Revolutions	Compensation	K_{lim}	
			Theoretical	Experimental
Analog	None	None	∞	---
Mixed	13 D. R.	None	8.2	8.2
Mixed	13 D. R.	$W_1^*(z)$ Eq. (18)	27.4	26.8

Conclusion

This paper has reported on particular systems whose performance was improved by suitable programs in the digital computer. It was assumed that justification for using a digital computer as part of the closed-loop system was based on several considerations, only one of which is the compensating potentiality of the computer. By discussing both some analytical and experimental means for the design of programs for system compensation it is hoped that the state of the art has been furthered. The demonstration of the extension of conventional techniques of frequency analysis to deal with the problem of sampling and sampled-data compensation has been successful, as close agreement between analytical and experimental results shows. The systems studied were purposefully chosen to approach their dynamic limitations.

Only the highlights of the experiment are discussed in this report. Further details of the experimental setup and the theoretical background as applicable to these experiments are given in References 6 and 7. Several other compensating programs were also developed analytically and experimentally with varying degrees of success. It is hoped that these studies lead to more systematic means for designing compensating programs of the general form. The effect of noise on the system was examined qualitatively, but this investigation was neither systematic enough nor were the results unusual to warrant its inclusion into this report.

The aid of Messrs. J. D. Cloud and T. J. Burns in this work is gratefully acknowledged.

References

1. M. L. MacKnight and P. A. Adamson, "Multichannel Analog Input-Output Conversion System for Digital Computer," Convention Record of the I.R.E. 1953 National Convention, Part 7, Electronic Computers.

2. W. K. Linvill, "Sampled-Data Systems Studied through Comparison of Sampling with Amplitude Modulation," Transactions of the A.I.E.E., Volume 70, Part II, pp. 1779-1788; 1951.

3. J. R. Ragazzini and L. A. Zadeh, "The Analysis of Sampled-Data Systems," Transactions of the A.I.E.E., Volume 71, Part II, pp. 225-234; 1952.

4. William K. Linvill and John M. Salzer, Analysis of Control Systems Involving Digital Computers," Proceedings of the I.R.E., Volume 41, No. 7; July, 1953.

5. John M. Salzer, "Frequency Analysis of Digital Computers Operating in Real Time," Proceedings of the I.R.E., Volume 41, No. 7; July, 1954.

6. T. J. Burns, J. D. Cloud and J. M. Salzer, "Experiments with a Digital Computer in a Simulated System," Technical Memorandum, TM-338, Hughes Research and Development Laboratories, Culver City, California; 1954.

7. T. J. Burns, J. D. Cloud and J. M. Salzer, "Experiments with a Digital Computer in a Simple Control System," paper presented at the Western Computer Conference, February 11, 1954.

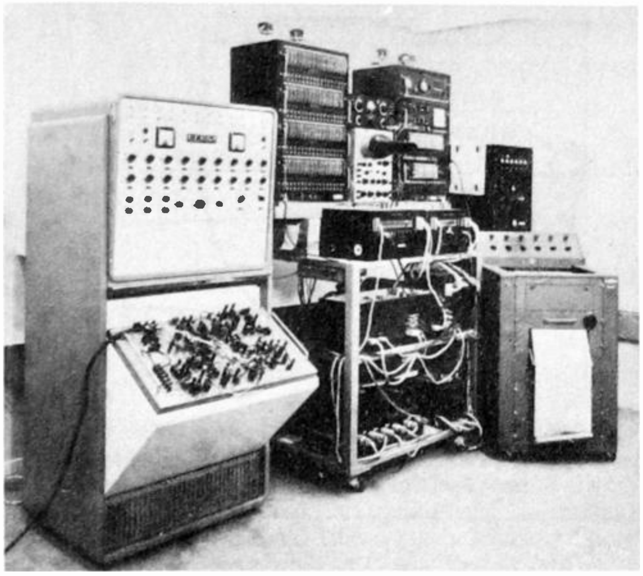


Fig. 1
Photograph of experimental setup.

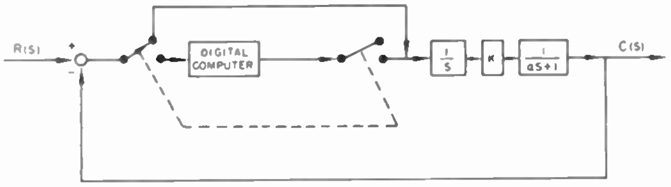


Fig. 2
Simulator setup for analog and mixed system tests.

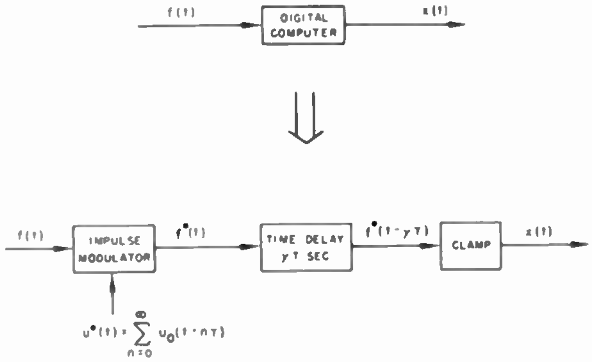


Fig. 3
Equivalent circuit of digital computer.

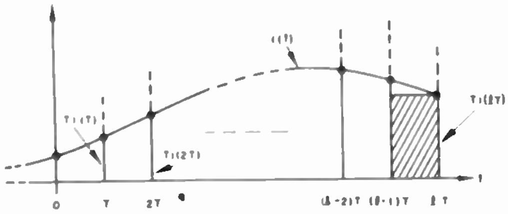
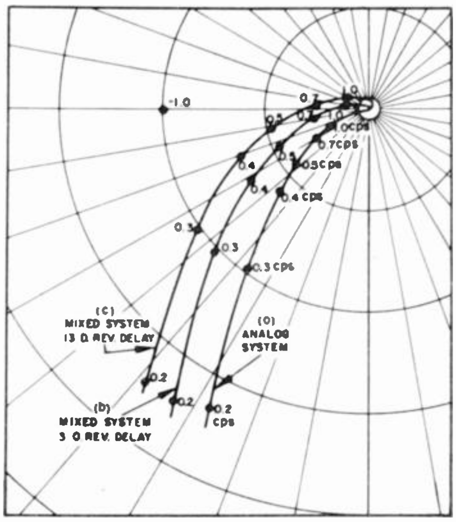


Fig. 4
Approximation of a continuous function.

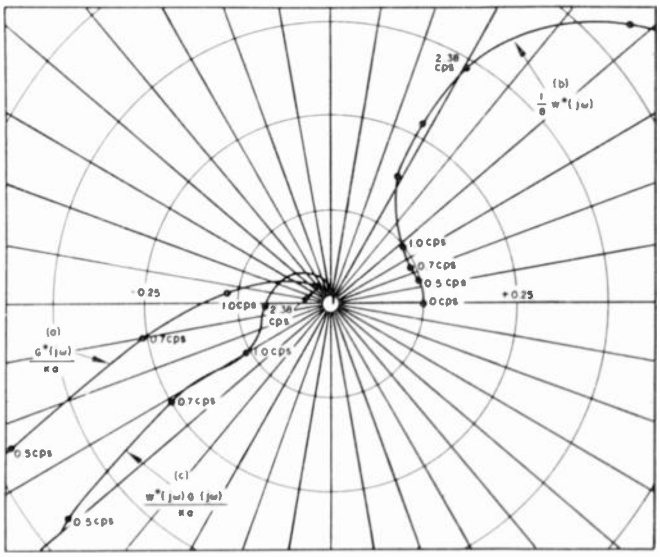


(a) $G(S)/Ka = G_2(S)/Ka = \dfrac{1/0.42}{S(0.42S+1)}$

(b) $G^*(S)/Ka \approx G(S)/Ka \cdot e^{-\gamma TS} \cdot \dfrac{1-e^{-TS}}{S} \cdot \dfrac{1/0.42}{S(0.42S+1)}$; $\gamma = 3/14$, $T = 1.05$ SEC.

(c) $G^*(S)/Ka$ SAME AS (b) BUT $\gamma = 13/14$ T I.

Fig. 5
Nyquist plots for narrow-band case without compensation.



(a) $G^*(S)/Ka \approx G(S)/Ka \cdot e^{-3TS/14} \dfrac{1-e^{-TS}}{S} \cdot \dfrac{1/0.42}{S(0.42S+1)}$; $T = 0.105$ SEC.

(b) $\dfrac{1}{8} w^*(S) = \dfrac{1}{8}\left[2.5 - 2.5e^{-TS} + e^{-2TS}\right]$

(c) $w^*(S) \cdot G^*(S)/Ka \approx w^*(S)G(S)/Ka$

Fig. 6
Nyquist plots for narrow-band case with compensation.

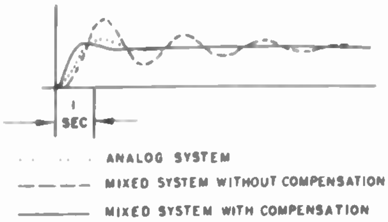
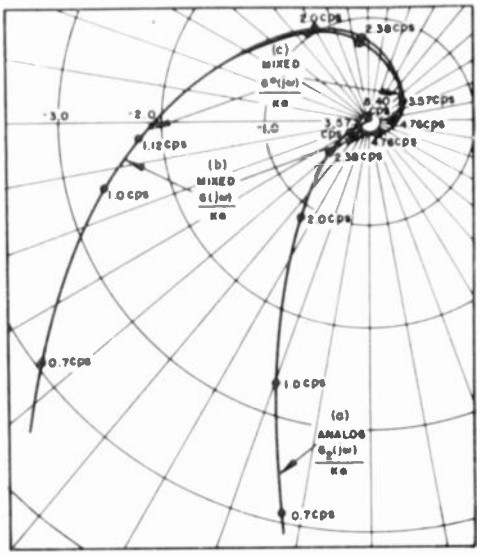


Fig. 7
Step response of narrow-band system.



(a) $G(s)/K_a \cdot G_2(s)/K_a = \frac{1/0.056}{s(0.056s+1)}$

(b) $G(s)/K_a \cong e^{-Ts} \cdot \frac{1-e^{-Ts}}{s} \cdot \frac{1/0.056}{s(0.056s+1)}$; $T = 0.105$ SEC.

(c) $G^*(s)/K_a \cong e^{-Ts} \cdot \left[3.67 + \frac{1.88}{1-e^{-Ts}} - \frac{5.55}{1-0.153e^{-Ts}}\right]$; $T = 0.105$ SEC.

Fig. 8
Nyquist plots for wide-band case without compensation.

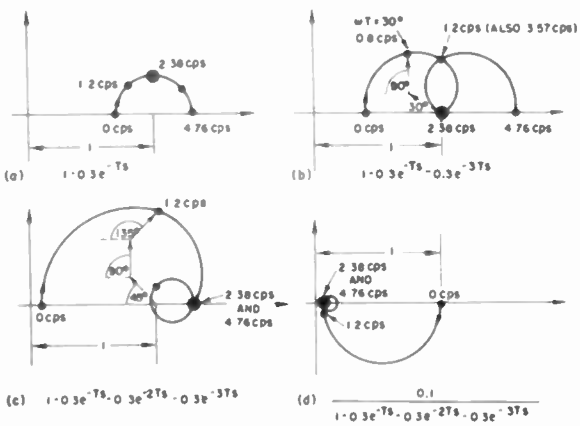


Fig. 9
Development of initial compensating program for wide-band system.

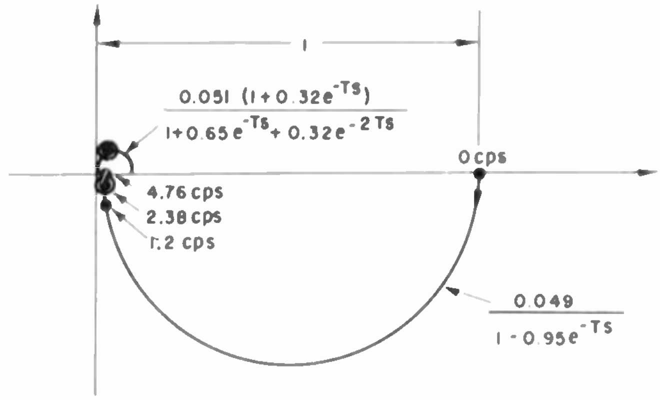
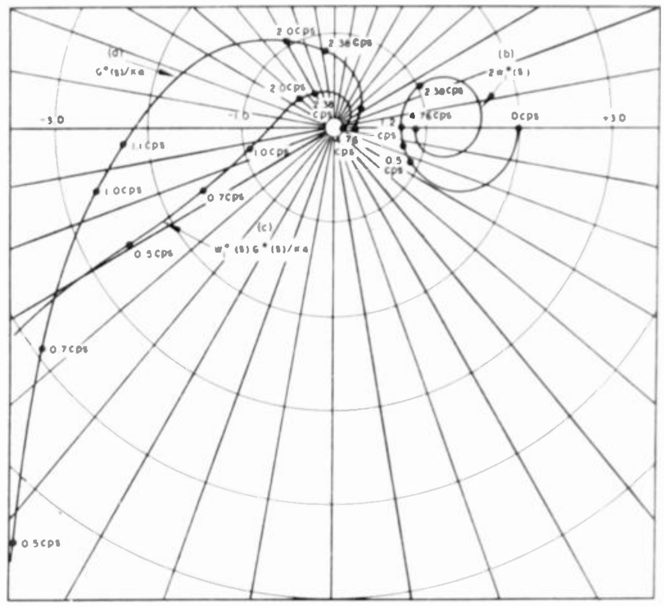


Fig. 10
The partial fractions of the initial compensating program for wide-band system.



(a) $G^*(s)/K_a \cong e^{-Ts}\left[3.67 + \frac{1.88}{1-e^{-Ts}} - \frac{5.55}{1-0.153e^{-Ts}}\right]$; $T = 0.105$ SEC

(b) $2W_1^*(s) = 2\left[\frac{0.075}{1-0.895e^{-Ts}} + \frac{0.452(1+0.33e^{-Ts})}{1+0.7e^{-Ts}+0.4e^{-2Ts}}\right]$, $T = 0.105$ SEC

(c) $W_1^*(s)\,G^*(s)/K_a$

Fig. 11
Nyquist plots for wide-band case with compensation.

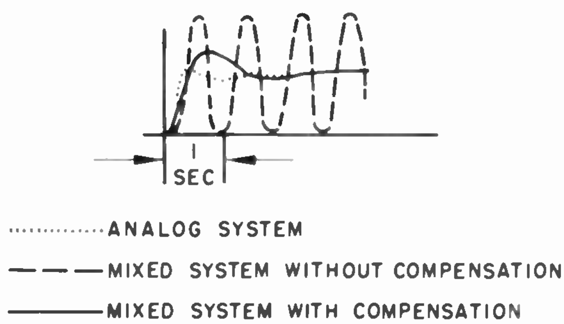


Fig. 12
Step response of wide-band system.

BINARY CONTROL SYSTEM FOR DIGITAL-TO-SHAFT POSITION MECHANISMS

by

ARTHUR H. WULFSBERG
ENGINEERING AND RESEARCH DIVISION
COLLINS RADIO COMPANY
CEDAR RAPIDS, IOWA

Summary

A system for the control of rotary actuators or indicators from binary coded information is described. By means of this system, mechanisms having considerable output torque may be controlled by means of punched cards or other forms of binary registers. A simple 8-digit 256-position digital-to-analogue mechanism is described. The control system has also been found useful in performing the remote frequency selection function in communications and navigation equipment.

The use of binary numbers for handling information of all types has become very common, primarily because of the many means available for generating, handling and storing binary digits. From the electron tube to the toggle switch, there are any number of electrical and electronic devices which like to say "yes" or "no" rather than "how much".

In certain types of control and indicating systems, it is desirable to set the position of a rotary shaft in accordance with a binary number which may be the output of a computer or a data transmission system. Such mechanisms are usually referred to as "digital-to-analogue transducers".

It is the purpose of this paper to describe a rather simple mechanism of this type: specifically, a device capable of setting a rotary shaft to any one of 256 discrete positions in accordance with an 8-digit binary number. This device is not particularly rapid in operation and only unidirectional rotation is provided. Its advantages include simplicity, mechanical accuracy and the ability to provide relatively high output torque.

Also to be described is a variation of this system which has been widely employed in remote control systems for multi-frequency radio communications and navigation equipment. While these systems are not concerned with binary numbers as such, the control information is handled in binary or on-off form. This fact has been of great value in these applications because of the ease and reliability of handling control information by means of simple switches and relays. Also, the control information can be stored conveniently by means of punched cards or similar memory devices.

In the systems to be described, the mechanism employed for turning the shaft is the Collins Autopositioner, a simplified diagram of which is shown in Figure 1. The basic elements of this device consist of a motor and its speed-reduction gear train, a slip clutch, a rotating shaft to which is fastened a notched stop-wheel, a pawl which engages the notches in the stop-wheel and a relay which actuates the pawl and operates a pair of electrical contacts for starting and stopping the motor.

The control circuit shown is a simple "open-seeking" system in which one control wire is needed for each possible position of the shaft. When the control switch is set to a given position, it grounds the corresponding control wire, closing the circuit from the relay to ground. The relay lifts the pawl from the stop-wheel notch and causes the motor to operate, driving the shaft and the seeking switch rotor until the proper position is reached, at which time the relay circuit is opened, the pawl drops into the stop-wheel notch and the motor stops. The slip clutch between the motor and the controlled shaft absorbs the energy of the motor as it coasts to a stop and also permits the operation of two or more Autopositioner units from a common motor. In this case the motor control contacts of the Autopositioner relays are connected in parallel so that the motor will operate so long as any of the relays are actuated.

The accuracy of this device is determined principally by the accuracy with which the notches are milled into the periphery of the stop wheel. Positioning accuracy is normally held to about ±1/3 degree; reset accuracy is approximately ±0.05 degree. In the size of unit used to date, output torque from 1 to 6 inch pounds can be provided. While there are many types of mechanisms available for performing the function of shaft rotation, this device has been found to be very adaptable to use in radio equipment because of its simplicity, accuracy, and the fact that its motion is continuous rather than intermittent. This last factor is of considerable importance when the load to be driven has appreciable inertia. In addition, a wide range of speed, torque and positioning accuracy can be obtained by the use of suitable gearing between the mechanism and the driven load.

Figure 2 is a diagram of the control system used for setting the mechanism in accordance with

a binary number. The circuit shown is for a 4-digit number, giving 16 different combinations, or decimal numbers from zero to 15. This system employs one control switch and one "seeking" switch for each digit of the binary number. With the control and seeking switches set symmetrically as shown, there is no circuit from the relay to ground. If the settings of the control and seeking switches are different from each other, a circuit to ground will be made, causing the mechanism to operate until symmetry is restored. For example, if switch 1 is set to "B", a circuit through S-2 to ground will be maintained until S-2 is also set to "B". If S-7 is set to "A", a circuit through the common wire and S-8 to ground will be maintained until S-8 is also set to "A". Note that the wire connecting all the "A" terminals of the switches is in parallel with those control wires whose switches are set to "A". It is actually needed in the circuit only when all the switches are set to "B".

This system could be made physically realizable for the control of a rotating shaft by using for the seeking switch a group of leaf-type single-pole-double-throw switches actuated by coded cams mounted on the rotating shaft to turn the switches on and off as the shaft rotates.

To eliminate the bulk and complexity of such a system, special wafer switches with their rotor blades cut in accordance with the binary code are used. The diagram of Figure 3 shows the complete circuit for a 16-position 4-digit control system using these switches. The electrical circuit is identical to that shown in Figure 2. As the output shaft rotates clockwise, the switch rotors will set up connections according to the familiar binary code as illustrated in the chart. The system shown can be expanded to handle any number of binary digits within physical limitations.

A certain application required an 8-digit 256-position mechanism. This was constructed from two 16-position mechanisms using the circuit shown. The two mechanisms were coupled to the output shaft through a planetary type differential mechanism which added the motion of the two shafts to produce an output shaft having 256 positions in 360 degrees of rotation.

Figures 4 and 5 are photographs of a model which was made up primarily for demonstration purposes. The mechanism at the right is controlled by the 1, 2, 4, and 8 value control switches and its dial reads from zero to 15. The mechanism at the left is controlled by the 16, 32, 64 and 128 value switches, and its dial reads zero, 16, 32, 48 and so forth, up to 240. The center dial is mounted on the output shaft and reads from zero to 255.

A simpler and more compact model of the 256-position mechanism is shown in Figures 6 and 7. This model also employs two 16-position mechanisms but the method of adding their motion is much simpler than that used in the demonstrator. The adding function was provided by mounting the "coarse" mechanism--that is, the one controlled by the 16, 32, 64 and 128 digits--on a pivot and rotating the entire mechanism over a small arc by means of a cam driven by the "fine" mechanism. The "fine" mechanism thus provides 16 discrete positions of the output shaft for each of the 16 positions of the "coarse" mechanism, or a total of 256 positions. This particular mechanism is controlled by 8 armature-type relays actuated by electron tubes. The output shaft drives an indicating dial and a synchro transmitter for transmitting the shaft position to a remote indicator.

As pointed out earlier, this mechanism does not have sense--that is, it does not go to its new position by the shortest possible path. The control circuit itself does not provide sensing information, so the mechanisms are merely driven always in the same direction until the new position is reached. The primary virtues of this device are simplicity and reasonably good accuracy. The time required to change to a new position varies from about one-tenth second to 1.5 seconds.

This mechanism, operating in direct accord with the familiar binary code, has not seen a great deal of application, but a variation of it has been employed in a number of airborne and ground station communications and navigation equipments for providing remote selection of operating frequencies. In such systems, it is desirable to provide a positive and reliable control system having a minimum number of interconnecting wires. In early multi-frequency airborne communication equipment, such as the World War II ART-13 HF transmitter and ARC-1 VHF transmitter-receiver, the open seeking system shown in Figure 1 was used.

This system requires one control wire for each operating frequency. In the case of the 10-channel equipments mentioned, 10 wires had to run from the cockpit frequency selector to the equipment rack.

With the advent of commercial use of VHF communications and navigation systems after the war, receivers and transmitters providing hundreds of frequencies selectable from the cockpit were needed. To minimize the number of control wires required, a system similar to the binary code system was developed. The circuit used for this system is identical to that shown in Figure 2 except that the wire connecting the "A" terminals of all the switches is deleted. This system works in the same fashion as the other except that the combination in which all switches are set to "B" is not used because no circuit from the relay coil to ground can be provided when all the seeking switches are set to "B". This system is thus capable of providing $2^N - 1$ different positions, where N is the number of control wires employed. Thus, with 4 wires and a ground connection, 15 positions can be provided; with 5 wires, 31 positions, etc.

For use in controlling shaft rotating mechanisms, special control switches have been designed which set up various combination of connections as the rotors are turned. Figure 8 shows the circuit for a 4-wire 10-position control system. In this circuit, the relay coil is connected to the back rotor contact; the front rotor contact is connected to ground. As the rotor rotates, the front rotor blade will connect the control wires to ground in the sequence shown in the chart. All control wires not grounded are connected together and to the relay coil by the back rotor blade. The code has no particular significance--the only requirements for the switch being that no combination repeats itself and that the one unusable combination is not set up. The seeking switches must be of the shorting or make-before-break type so that the circuit will not be broken before the correct position is reached. Figure 9 shows a number of different control switches which have been used in this system; these switches provide from 10 to 24 positions. Standard phenolic and ceramic switch wafers have been used for this purpose.

Typical of the equipments employing this type of control system is the Collins Type 51R airborne VHF receiver (Figure 10). This receiver can operate on any one of 280 crystal controlled frequencies spaced at tenth-megacycle intervals from 108 to 136 megacycles to provide omni-range navigation, instrument landing system and voice communication service for all types of aircraft. Originally developed for use in commercial transport aircraft, a military model designated AN/ARN-14 has been built in large quantities for the Air Force. In this receiver, two Autopositioner mechanisms are employed to select the frequency-determining crystal units and to tune the various amplifier circuits for operation at the frequency selected in the cockpit. One mechanism operates a 20-position shaft; the other a 14-position shaft, giving 280 different operating frequencies.

The cockpit frequency selector (Figure 11) includes two coaxial control switches, the dials of which are arranged to indicate operating frequency directly in megacycles and tenth megacycles. Nine control wires are used to provide the frequency selection function.

This type of control system has been used in a wide variety of commercial and military multi-frequency radio equipments. Aeronautical Radio, Inc., which is the standardizing agency for the scheduled airlines, has specified this system for use on several new VHF and high-frequency airborne communications transmitters and receivers.

While this system was developed primarily to minimize the number of control wires used, the fact that the control information is handled in binary or on-off form has led to several interesting applications of the system.

Figure 12 shows a punched-card reader for controlling the frequency of the airborne navigation receiver previously described and to supply data to a navigation computer.[1] The frequencies of the various navigation aids to be used on a flight can be punched into the card in the sequence in which they are to be used. As the flight progresses, the pilot merely turns the knob to switch to the next navigation facility to be used. The holes in the punched card actuate small snap-action switches which are connected per the circuit of Figure 8.

The system is also useful in application in which it is desirable to control the position of a rotating shaft by means of multiple audio or ultrasonic tones. In this case, tone-operated relays serve as the control switches. Such a system has been used in a military UHF communications system to provide remote selection of ten operating frequencies over telephone wires or radio links.

In conclusion, this type of control system has been found useful in a wide variety of applications and can be used with a number of different types of actuating mechanisms. In addition to the Autopositioner mechanism described, stepping switches, rotary solenoids and motors provided with braking facilities can be used where their characteristics are suitable for the application.

Acknowledgment must be given to Mr. H. M. Schweighofer of the Collins Radio Company, who is primarily responsible for the development of this system.

---

1 E. H. Fritze, "Punched-Card Controlled Aircraft Navigation Computer", Proc. I.R.E. Vol. 41, No. 6, June 1953.

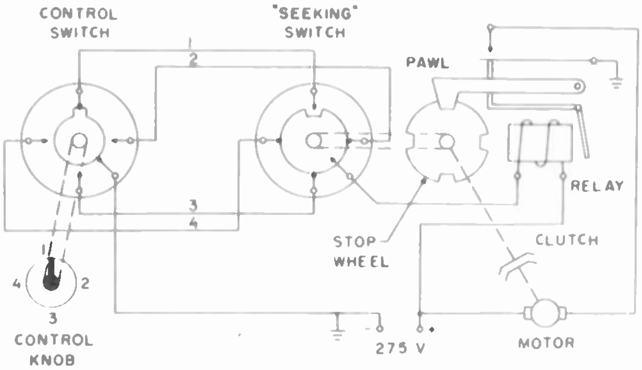


Fig. 1
Simplified diagram - autopositioner mechanism.

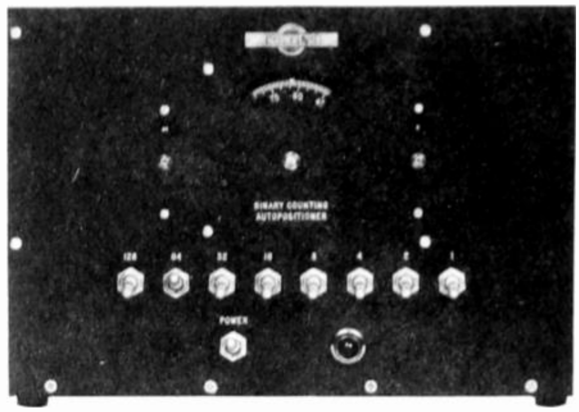


Fig. 4
Binary autopositioner model - front view.

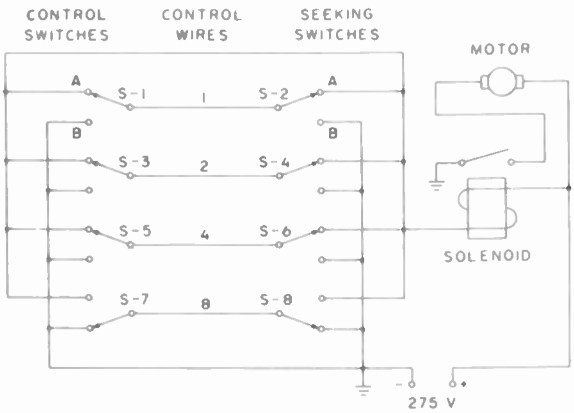


Fig. 2
Binary control system circuit.

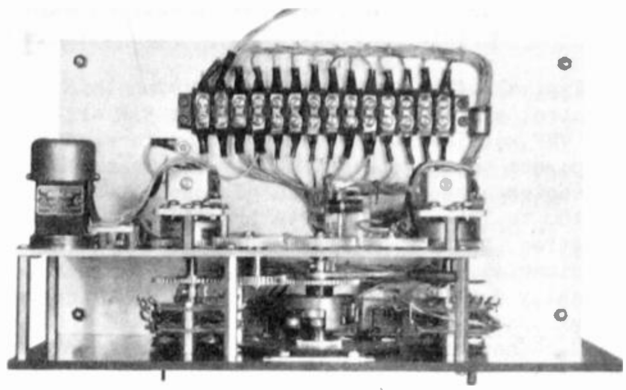
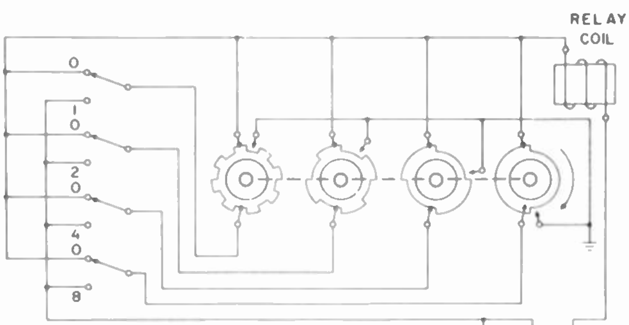


Fig. 5
Binary autopositioner model - top view.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		x		x		x		x		x		x		x		x
2			x	x			x	x			x	x			x	x
4					x	x	x	x					x	x	x	x
8									x	x	x	x	x	x	x	x

Fig. 3
Binary control system circuit showing coding switches.

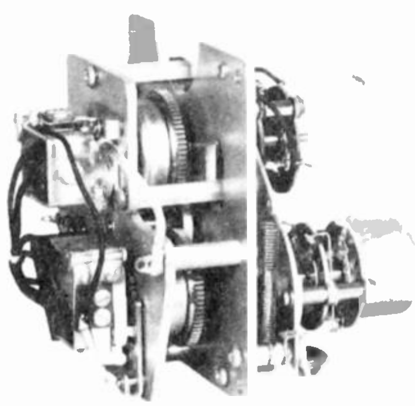


Fig. 6
Binary autopositioner mechanism.

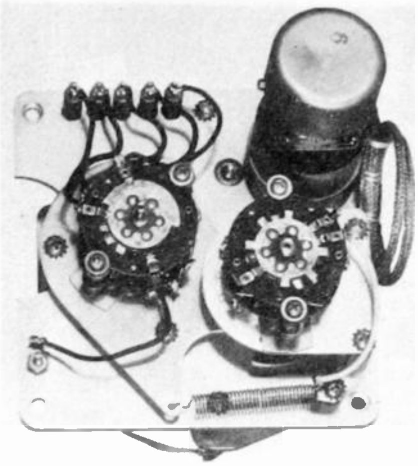
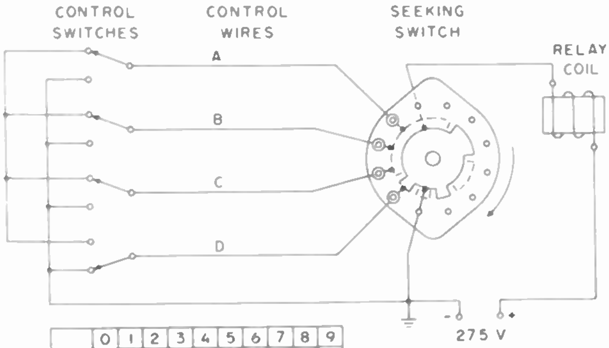


Fig. 7
Binary autopositioner mechanism.



	0	1	2	3	4	5	6	7	8	9
A				x		x			x	x
B			x		x			x	x	
C		x		x			x	x		
D	x		x			x	x			

Fig. 8
10-positon control circuit.

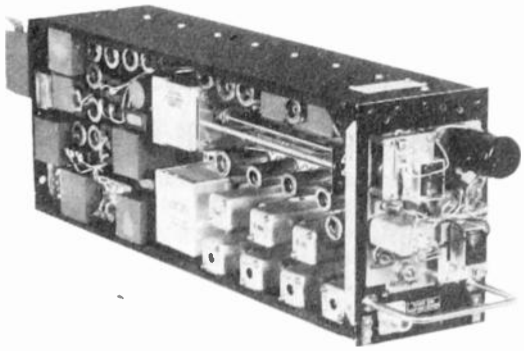


Fig. 10
Collins type 51R vhf navigation receiver.

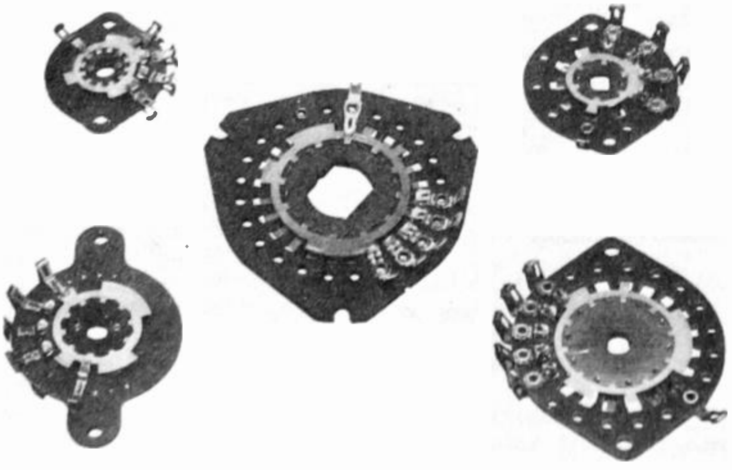


Fig. 9
Typical control switches.

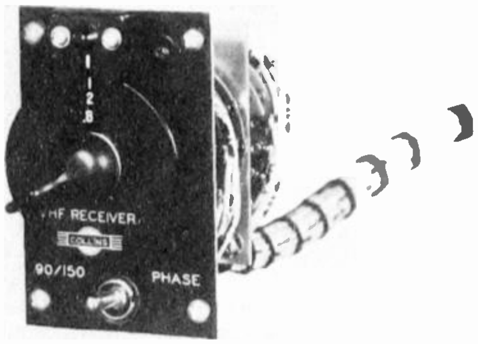


Fig. 11
Frequency selector for navigation receiver.

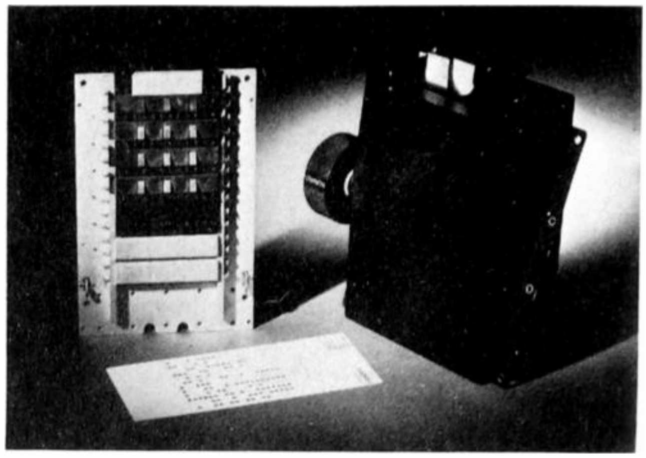


Fig. 12
Punched card control unit for navigation receivers and computer.

OPTIMIZATION OF SERVOSYSTEMS

Richard C. Lyman and William P. Caywood, Jr.
Carnegie Institute of Technology
Pittsburgh, Pennsylvania

Summary

An analytical method is presented for the design of optimum servosystems which are subjected to signal inputs contaminated with noise. Minimum mean square error is the criterion of optimization. A mean square error spectrum is derived, using a probability theory operation termed Expected Value. The method is applied to several examples of random behavior in the signal and noise spectra and in the parameters of the system function.

Introduction

This paper is a description of an analytical method for the design of optimum servosystems which are subjected to signal inputs contaminated with noise. The method is especially applicable to those design problems characterized by parameters which exhibit time-varying, random behavior. It can also be applied to most general time-invariant problems as special cases.

This method is primarily carried out in the frequency-plane rather than in the time-plane. Hence the concept of the frequency components present in the system inputs, output, and error, as described by their spectral densities, is used rather than the concept of their variation with respect to time.

The servosystem to be optimized is described by a general transfer function composed of the system attenuation and phase shift as functions of frequency.

The method is limited to analysis and synthesis of linear systems. It uses minimum mean square error as the criterion of system optimization.

A concept which is basic to this use of spectral densities stems from its being a statistical approach to the optimization problem. The system input is generally not known with certainty at any instant of time or as a function of time, but rather it may be described only by its statistical properties. These properties, which may be spectral densities or auto-correlation functions, must be obtained by some average of the inputs expected over the entire period of operation for which the system is being designed, or over a sufficient statistical sample of that period. Such information will yield only the statistical properties of the error such as its mean square spectral density, as is used in this method. As a result of this statistical approach, the method optimizes the overall performance of the system rather than any particular portions of it.

Derivation of the Squared Error Spectrum

The description of this method begins with the derivation of the mean square error spectrum. For the purposes of this derivation, the signal and noise inputs to the system are represented by voltage spectral densities rather than by power spectral densities. A voltage spectral density may be considered to be composed of an amplitude spectrum and a phase spectrum, thus indicating the relative potential of each frequency component present and the phase angle of each component. Both spectra may be determined using the Fourier transform if the signal may be described functionally in time. Otherwise, actual measurement using a wave analyser or like instrument can yield the amplitude spectrum only. Primarily because of this limitation, only the amplitude spectra are used in this method.

The input to the system to be optimized is assumed to consist of some signal contaminated with noise. A particular frequency component of the signal in the time-plane may be written as the real part of a vector rotating at that frequency ω_1 and having for its magnitude the value of the signal amplitude spectrum $\mathcal{S}(\omega_1)$:

$$\mathcal{R}e\left[\mathcal{S}(\omega_1)\,\varepsilon^{j\omega_1 t}\right] \qquad (1)$$

The noise may be represented similarly, with the addition of a relative phase angle ϕ which refers the absolute phase angle of the noise to that of the signal:

$$\mathcal{R}e\left[\mathcal{N}(\omega_1)\,\varepsilon^{j\omega_1 t+j\phi}\right] \qquad (2)$$

The sum of these vectors is the system input, consisting of signal contaminated with noise:

$$\mathcal{R}e\left[\mathcal{S}(\omega_1)\,\varepsilon^{j\omega_1 t} + \mathcal{N}(\omega_1)\,\varepsilon^{j\omega_1 t+j\phi}\right] \qquad (3)$$

The system output spectral density is the result of this input operating on the system function, denoted by the familiar: $M(\omega_1)\,\varepsilon^{j\alpha(\omega)}$, where $M(\omega)$ is the attenuation and $\alpha(\omega)$ is the phase shift. A particular frequency component of this output spectrum is written as:

$$\mathcal{R}e\left[\left\{\mathcal{S}(\omega_1)\,\varepsilon^{j\omega_1 t} + \mathcal{N}(\omega_1)\,\varepsilon^{j\omega_1 t+j\phi}\right\} M(\omega_1)\,\varepsilon^{j\alpha(\omega_1)}\right]$$

which will be abbreviated as:

$$\mathcal{R}e\left[\{\mathcal{S}\,\varepsilon^{j\omega_1 t}+\eta\,\varepsilon^{j\omega_1 t+j\phi}\}M\varepsilon^{j\alpha}\right] \qquad (4)$$

The same frequency component of the error is the difference between the signal and output components, equations (1) and (4). The real part of this frequency component of the error is:

$$e(t;\omega_1)=\mathcal{S}\,\mathrm{Cos}(\omega_1 t)-\mathcal{S}M\,\mathrm{Cos}(\omega_1 t+\alpha)-\eta M\,\mathrm{Cos}(\omega_1 t+\alpha+\phi). \qquad (5)$$

In order to form the ω_1 frequency component of the mean square error spectrum, the error function, equation (5), must be squared and averaged first over one cycle of its sinusoidal time variation, holding the input spectra and system parameters temporarily constant. That is:

$$e^2(\omega_1)=\frac{\omega_1}{2\pi}\int_0^{\omega_1/2\pi}\{e(t;\omega_1)\}^2\,dt$$

$$e^2(\omega_1)=\tfrac{1}{2}\{\mathcal{S}^2+(\mathcal{S}^2+\eta^2)M^2-2\mathcal{S}^2M\,\mathrm{Cos}(\alpha)-2\mathcal{S}\eta M\,\mathrm{Cos}(\alpha+\phi)+2\mathcal{S}\eta M^2\,\mathrm{Cos}(\phi)\} \qquad (6)$$

Because $\mathcal{S}(\omega)$ and $\eta(\omega)$ were considered to be the magnitudes of the input signal and noise spectra, they are peak values. However, amplitude spectra are commonly given in rms values, as could be obtained by taking the square root of the power spectra. Therefore, the symbols S and N will henceforth be used to represent the rms values of the signal and noise spectra. As a result of this change from peak values to rms values, the factor of $\frac{1}{2}$ multiplying equation (6) is removed, and the squared error spectrum is written as:

$$e^2(\omega)=S^2+(S^2+N^2)M^2-2S^2M\,\mathrm{Cos}(\alpha)-2SNM\,\mathrm{Cos}(\alpha+\phi)+2SNM^2\,\mathrm{Cos}(\phi) \qquad (7)$$

This squared error spectrum is a function of the parameters describing the signal and noise input spectra (S, N, and ϕ) and the system function (M and α). If any or all of these parameters exhibit some random behavior, the spectrum must be averaged over all of their possible values to form the mean square error spectrum. It is this latter spectrum that is to be minimized by proper choice of the system attenuation (M) and phase shift (α). This process of averaging over the random values of the parameters will be accomplished by the use of a probability theory operation termed Expected Value or Mathematical Expectation.

The Expected Value of a function of any number of random variables is defined[1] as the average value of that function when its random variables take on all of their permissible values with their associated probabilities of occurrence. Briefly, the Expected Value of a function $f(x)$ of the random variable x, where x takes on each value with a probability $p(x)$, is denoted by $E[f(x)]$ and is determined mathematically by either (1) summation:

$$E[f(x)]=\sum_{\text{all }x}f(x)\,p(x)$$

if x takes on only discrete values in which case $p(x)$ is termed the discrete probability density function of x, or (2) integration:

$$E[f(x)]=\int_{\text{all }x}f(x)\,p(x)\,dx$$

if x varies continuously in which case $p(x)$ is termed the probability density function of x. For the functions treated in this application, the Expected Value of a sum is the sum of the Expected Values, the Expected Value of a constant is that constant, and the Expected Value of a product of functions of statistically independent random variables is the product of the Expected Values. A property of the probability density function $p(x)$ is that

$$\int_{\text{all }x}p(x)\,dx=1$$

since it is certain that x will take on some value within its range of permissible values.

Application of the Mean Square Error Spectrum

Introduction

A type of problem to be solved by the method of this paper is the determination of the system function that minimizes the mean square error spectrum. That spectrum may be written as the Expected Value of equation (7):

$$\overline{e^2}(\omega)=E\left[e^2(\omega)\right], \qquad (8)$$

The error spectrum is a function of the input spectra and of the system function, both of which are functions of frequency in general. Also, both may be permitted to exhibit some degree of random behavior which may possibly be dependent in the probability sense on the behavior of the other parameters. Because of the many possibilities for such varied and interrelated random behavior, various sets of restrictive assumptions will be made to facilitate description of some applications of this method.

Assumption of Random Noise

For all applications treated in this paper, the random behavior of the noise input is assumed to be independent of that of any other parameter (eg. S, M, or α). The relative phase angle ϕ, which relates the absolute phase angles of noise and signal, is assumed to be uniformly random, taking on all values with equal probabilities. It is also assumed to be independent of the other parameters (S, N, M, and α), in the probability sense. It results that the Expected Value of $\mathrm{Cos}(\phi)$ is zero and the last two terms of the expression for the mean square error spectrum are zero. Hence equations (7) and (8) become:

$$\overline{e^2}(\omega)=E\left[S^2+(S^2+N^2)M^2-2S^2M\,\mathrm{Cos}(\alpha)\right]. \qquad (9)$$

Invariant Spectra and System

The set of additional assumptions for this first, introductory example are very restrictive. However, it illustrates by a new method an application and a result previously described in the literature.[2,3] The system attenuation and phase shift are restricted to be time-invariant. The signal and noise spectra are assumed to remain

invariant over the period of operation for which the system is being designed. Since all the parameters of the mean square error spectrum are assumed to be constant, the Expected Value operation on equation (9) yields:

$$\overline{e^2}(\omega) = S^2 + (S^2+N^2)\,M^2 - 2\,S^2 M \cos(\alpha). \qquad (10)$$

The synthesis problem becomes that of determining the system attenuation and phase shift as functions of the signal and noise parameters to minimize the mean square error spectrum.

The phase shift ($\underline{\alpha}$) at a particular frequency is a unique function of the attenuation ($\underline{M}$) at all frequencies for nearly all servosystems (ie. for minimum phase, linear, stable systems composed of fixed, lumped parameters) as expressed by Bode and others.[4] The derivative of the mean square error spectrum with respect to $\underline{M}$ could be equated to zero to determine the optimum $\underline{M}$ from which the optimum $\underline{\alpha}$ might be calculated. Since the relationship between $\underline{\alpha}$ and $\underline{M}$ is so intractable as to prevent substitution, the classical assumption of $\underline{\alpha}$ independent of $\underline{M}$ will be made. Independence allows the partial derivatives of the spectrum with respect to $\underline{\alpha}$ and $\underline{M}$ to be each equated to zero and solved for their optimum values.

$$\delta\overline{e^2}(\omega)/\delta\alpha = 2S^2 \sin(\alpha) = 0\ ,$$

which requires that $\underline{\alpha} = 0$; that is, a zero phase shift system is indicated to be optimum.

$$\delta\overline{e^2}(\omega)/\delta M = 2M(S^2 + N^2) - 2S^2 \cos(\alpha) = 0\ ,$$

which yields, with $\underline{\alpha} = 0$:

$$M(\omega) = S^2(\omega)/\{S^2(\omega) + N^2(\omega)\} \qquad (11)$$

Wiener[5] develops the same result in an effectually different manner.

Time-varying Spectra

The shapes of the signal and noise spectra describing the inputs to the system may be permitted to vary in time by another, different set of restrictions. The signal spectra are assumed to vary independently of the noise spectra in the probability sense. The assumption of uniform random behavior of the relative phase angle ϕ is retained. Under the conditions stated the mean square error spectrum becomes:

$$\overline{e^2}(\omega) = E[S^2] + M^2 E[S^2] + M^2 E[N^2] - 2M \cos(\alpha) E[S^2]. \qquad (12)$$

Use of the expression for the mean square error spectrum, equation (12), requires the mean square value of the signal amplitude spectra and of the noise amplitude spectra. The mean square spectrum of a set of amplitude spectra is equivalent to the average spectrum of the power spectra. It is apparent that, because the mean square criterion is essentially a power criterion, the use of the average spectrum of the power spectra available to describe the signal or noise constitutes the best use of the information available. No additional information is required or can be used. The average is weighted according to the relative durations or probabilities of occurence of the various spectra, as the Expected Value operation indicates.

If $\underline{E[S^2]}$ and $\underline{E[N^2]}$ are denoted by $\underline{P_S}$ and $\underline{P_N}$, which are averages of the respective sets of power spectra available, the ideal transfer function may be derived in the same manner as was done in the previous example. It will have zero phase shift and an attenuation of:

$$M(\omega) = P_S \,/\, (P_S + P_N)\ .$$ [5]

Formulation of the Mean Square Error

Thus far, application of this method has been restricted to systems having zero phase shift. Recognizing the general impossibility of attenuation without phase shift in physically realizable and stable systems, the method is extended by integrating the mean square error spectrum over all real frequencies to form the mean square error. The problem of optimization becomes that of minimizing the mean square error. The most advantageous use of the extended, more realistic procedure lies in substitution of a particular type of system transfer function for the general function $\underline{M\epsilon^{j\alpha}}$ in the mean square error spectrum. Particular forms of signal and noise amplitude spectra are also substituted. The mean square error is minimized with respect to such system parameters as time constants, gain, damping ratio, and the like, within the limitations of system stability. There results the optimum values of the system parameters and the optimum characteristic of the type of system transfer function substituted.

The optimum system derived may be considered as that system which achieves the best compromise between passing the signal spectrum (and some noise along with it) and attenuating the noise spectrum (and some signal along with it).

Examples of Minimization

There is described in this part of the paper the application of the procedure just formulated to three general types of servosystems. The signal amplitude spectrum is assumed to be of first order form, as shown in figure 1. A white noise spectrum is assumed. The signal to noise ratio is described in terms of the values of the signal and noise amplitude spectra at zero frequency (dc). The ratio is denoted by $\underline{\Sigma}$ and is an important parameter in the examples to follow. The series of general servosystems to be optimized may be described by their open loop transfer functions, which are (figure 2): (1) an integrator, (2) a simple lag, and (3) a simple lag with an ideal amplifier.

For each example of the series, the input spectra shown in figure 1 and the system attenuation and phase shift are substituted in the proper form of the mean square error spectrum, see equation (10). The spectrum is integrated over all real frequencies, by the method of residues,

to form the mean square error. The values of the system parameters which minimize the mean square error for a given ratio of signal to noise are determined by ordinary or partial differentiation. These values are subject to the requirements of system stability. The parameters of these three systems are, for system (1) the integration constant τ, (2) the time constant τ, and (3) the gain K and time constant τ. Ordinary differentiation is used to optimize the first two, single-parameter systems. Partial differentiation is required to treat the third, a two-parameter system. In more complex examples of two-parameter systems, the partial differentiation becomes too intractable, and the surface generated by the mean square error as a function of the two parameters is better examined graphically for the minimum. The mathematics involved in this application is straightforward in essence. A few, simple, particular examples of such application have been carried out in the literature.[3]

The results of optimizing the three types of servosystems listed above are presented in figure 4. The mean square error of each optimized system is shown as a function of the signal to noise ratio Σ. The mean square error function for the zero phase shift system previously derived, equation (11), is included for comparison, despite the system's unrealizability, because it represents a lower bound on the mean square error attainable. These functions are normalized to the mean square error of a no-pass system (a system having no output). Figure 4 illustrates several important conclusions of interest for both system analysis and system synthesis. These are:

(1) A worthwhile integrator system cannot be designed for $\Sigma < 1$; all stable integrator systems have at least as much error as the no-pass system for any $\Sigma < 1$. The reason is that the integration constant yielding minimum mean square error, given by $\tau = 1/(\Sigma - 1)$, becomes infinite as Σ approaches unity, that is, the system bandwidth becomes zero. For $\Sigma < 1$, the optimum τ becomes negative, hence the system is unstable, and any positive τ yields more error than a no-pass system.

(2) The mean square error curve of the simple lag system is similar to that of the integrator. The optimum time constant is given by $\tau = 1/(\sqrt{3}\,\Sigma - 1)$ which becomes infinite when $\Sigma = 1/\sqrt{3}$. For smaller Σ the system is worthless. The system has an attenuation of $\frac{1}{2}$ at zero frequency. That this is desirable for Σ near unity is shown by its superiority to the integrator system which has an attenuation of unity at zero frequency. For Σ larger than 3.7 the integrator system is seen to be superior, indicating the desirability of unity attenuation at zero frequency for large signal to noise ratios.

(3) The simple lag with an ideal amplifier is seen to be superior to the first two systems almost everywhere. This superiority may be attributed to the gain parameter which permits changing the zero frequency attenuation, and hence the entire level of system attenuation, according to the ratio of signal to noise. At the point where $\Sigma = \sqrt{3}$, this system error coincides with that of system (2) because its optimum gain is unity. Below $\Sigma = \sqrt{3}$, its optimum gain is less than unity; the amplifier is only an attenuator. In this region of $\Sigma < \sqrt{3}$, system (2) could be improved easily by the addition of an attenuator in its open loop, making it the same as system (3).

(4) In addition to the systems described, a second order system was similarly treated. It has a closed loop transfer function composed of the product of two simple lag circuits with independent time constants. Optimization of this system with respect to the two time constants dictated that one of them be zero for minimum mean square error. Thus, under the assumptions made, the best second order system of this form is a simple first order system corresponding to the integrator system treated first.

Random Behavior of System Parameters

Introduction

In the preceeding examples, the servosystem was considered to be composed of fixed elements, that is of elements such as gain and time constants having constant values during the period of operation for which the system was being designed. It is possible that the environment of the system may vary greatly during that period. As a result of changes in pressure, temperature, humidity, and the like, the gain and time constants may vary appreciably from their intended values. Such variations are of more or less random nature.

For this example the system attenuation (M) and phase shift (a) are assumed to exhibit some random behavior independent of the signal and noise. The signal and noise spectra are assumed to be constant over the period of operation. However, the method could handle cases of dependence in the probability sense between random behavior in the system and time-variation in signal spectra.

The Form of the Mean Square Error Spectrum

Under the assumptions listed, the mean square error spectrum takes the form:

$$\overline{e^2}(\omega) = S^2 + (S^2 + N^2)E[M^2] - 2S^2E[M\cos(a)] \,. \quad (13)$$

The closed loop attenuation (M) and the phase shift (a) of the system are assumed to be functions of some system parameter τ which exhibits random behavior, having some average value and some probability density $p(\tau)$. The two Expected Values are generally evaluated by integration, for example:

$$E[M^2(\tau)] = \int_{\text{all } \tau} M^2(\tau)\cdot p(\tau)\cdot d\tau$$

If the system environment causes τ to take on any of several discrete values rather than vary continuously, or if continuous variation could be reasonably approximated by discrete values and a discrete probability density function, the integration is replaced by summation over all values of τ.

The Optimization Problem

The problem of designing the optimum system under these assumptions may be resolved into that of determining the best design value for the average of the random parameter. The general conclusion illustrated by this application of the method and by the example to follow is that a better system (having less mean square error) will result from this use of probabilities than would result if the random behavior of the parameter were neglected and its average value were considered as a nonrandom parameter. The improvement is a result of the fact that the mean square error expression contains system function averages, eg. $E[M^2(\tau)]$ and $E[M\pi\cos(a(\tau))]$, rather than averages of the parameter, eg. $E[\tau]$ or $E[\tau^2]$.

An Example

As an example, the servosystem composed of an integrator in the open loop, figure 2 - system (1), is considered. The time constant is assumed to exhibit, for this simplified example, a discrete random variation such that it may be either of two values, each with a probability of $\frac{1}{2}$. The two values may be written as $\tau_0 + \delta\tau_0$ and $\tau_0 - \delta\tau_0$ where τ_0 is the average value of the randomly varying time constant, and δ is the relative spread, which will be considered fixed. The design problem is the determination of the value of τ_0 which minimizes the mean square error.

Using white noise and the first order signal spectra which were used previously (figure 1), and the integrator system function as assumed above, the mean square error spectrum of equation (13) becomes:

$$\overline{e^2}(\omega) = \frac{\beta^2}{\omega^2+\alpha^2} + \left(\frac{\beta^2}{\omega^2+\alpha^2}+\eta^2\right)E\left[\frac{1}{1+\omega^2\tau^2}\right] - 2\frac{\beta^2}{\omega^2+\alpha^2}E\left[\frac{1}{1+\omega^2\tau^2}\right], \text{ where} \qquad (14)$$

$$E\left[\frac{1}{1+\omega^2\tau^2}\right] = \frac{1}{2}\,\frac{1}{1+\omega^2\tau_0^2(1+\delta)^2} + \frac{1}{2}\,\frac{1}{1+\omega^2\tau_0^2(1-\delta)^2}$$

The mean square error must be obtained by integration of this spectrum over all real frequencies. The operations of integration and Expected Value may be interchanged[6] under the conditions that $\overline{e^2}(\omega)$ is always less than some constant K and is integrable in the Riemann sense, and that $E[\overline{e^2}(\omega)]$ is also integrable in the Riemann sense. Integration by the method of residues yields:

$$\overline{e^2} = \frac{\pi}{2}\left\{\frac{\beta^2}{\alpha} + \eta^2 E\left[\frac{1}{\tau}\right] - \beta^2 E\left[\frac{1}{\alpha(1+\alpha\tau)}\right]\right\} \qquad (15)$$

which may be abbreviated as:

$$\overline{e^2} = E[\overline{e^2}(\tau)] = \tfrac{1}{2}\overline{e^2}(\tau_0+\delta\tau_0) + \tfrac{1}{2}\overline{e^2}(\tau_0-\delta\tau_0) \qquad (16)$$

The minimization of this expression with respect to the average value τ_0 of the time constant is accomplished by differentiation, subject to system stability considerations. However, a graphical presentation better illustrates the minimization process. A curve of the mean square error, equation (15), vs τ is presented in figure 3. As equation (16) indicates, the sum of the ordinates of that curve at $\tau_0 + \delta\tau = \tau$ and $\tau = \tau_0 - \delta\tau_0$ is to be minimized. Owing to the dissymmetry of the curve about its minimum, the optimum value of τ_0 will be shifted from the minimum to some larger τ, the amount of the shift depending upon the value of δ and the shape of the curve. For a particular example with $\Sigma = 2$ and $\delta = 0.5$, the optimum value of τ_0 is found to have shifted from the minimum of the curve at 1.0 to about 1.4. The mean square error for this τ_0 is 2% less than it would be for $\tau_0 = 1.0$. With other examples, the improvement may be significantly larger.

Other Applications

The use of probability concepts is demonstrated above to be necessary for the optimum design of those servosystems which are in essence similar to the example treated. The results of that example treated in detail above are of more general value than for the particular problem stated. As an illustration of this generality, the results can be interpreted to be the solutions of two different problems, which are described below.

(1) The results of the example can be used to indicate the greatest change that can be tolerated in a parameter's value without increasing the system mean square error appreciably. Such a change could result from thermal effects on a parameter.

(2) The problem of optimizing the average performances of a large number of servosystems which are the same except that their construction involves one or more elements having values within some manufacturing tolerance is merely a different interpretation of the example treated above. The average over the mean square errors of the systems containing such elements is equivalent to averaging the mean square error of one system as the values of those elements vary over their manufacturing distributions.

Conclusions

The examples described in this paper have indicated only a few of the possible sets of restrictions on the random behavior of system parameters and input spectra. The method presented here can be used to solve many problems of a less restricted nature. The treatment of cases

in whi h correlation exists between random variations in parameter values has not been described, but the solution of a few examples of that type has further substantiated the importance of the method.

References

1. Introduction to the Theory of Statistics, by A. McF. Mood, McGraw-Hill Book Company, 1950, Chapter 5.

2. "Separating Information from Noise", by Otto J. M. Smith, Trans. IRE Professional Group on Circuit Theory, December 1952, p. 83.

3. Theory of Servomechanisms, M.I.T. Radiation Laboratory Series, Volume 25, McGraw-Hill Book Company, 1947, Chapter 7, R. S. Phillips.

4. Servomechanisms and Regulating System Design, Volume I, by H. Chestnut and R. W. Mayer, John Wiley and Sons, Inc., 1951, pp. 297-302.

5. Extrapolation, Interpolation, and Smoothing of Stationary Time Series, by Norbert Wiener, Cambridge Technology Press of M.I.T., 1949, p. 95.

6. Foundations of the Theory of Probability, by A. N. Kolmogorov, Chelsea Publishing Co., 1950, Chapter 4, section 2 and 5.

This work was supported in part by The Office of Naval Research through contract number N7ori30308.

This paper is part of a dissertation submitted by Richard C. Lyman in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie Institute of Technology.

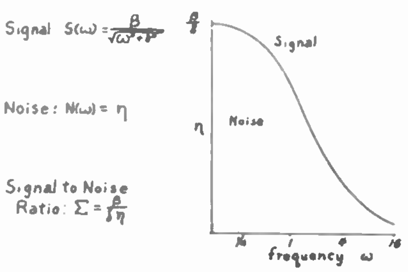


Fig. 1 - Input amplitude spectra.

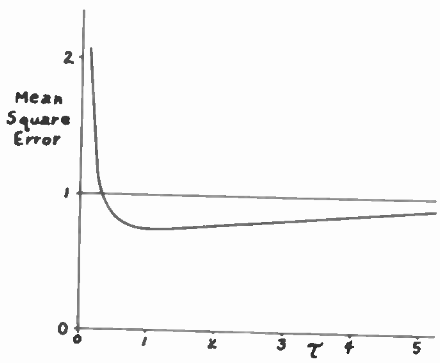


Fig. 3 - Integrator system.

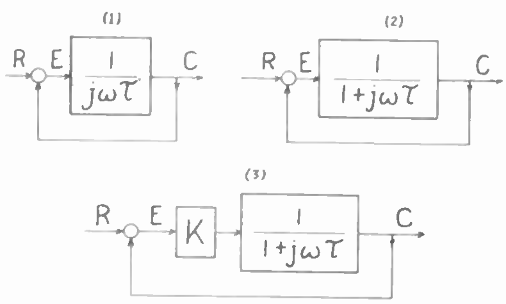


Fig. 2 - Servosystems.

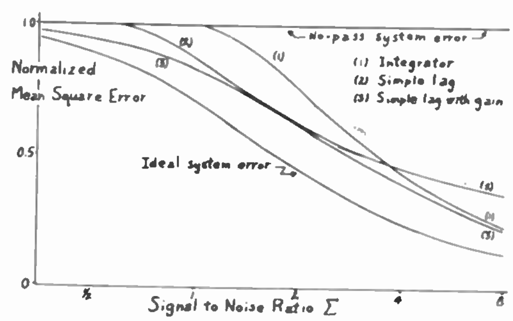


Fig. 4 - System mean square errors.